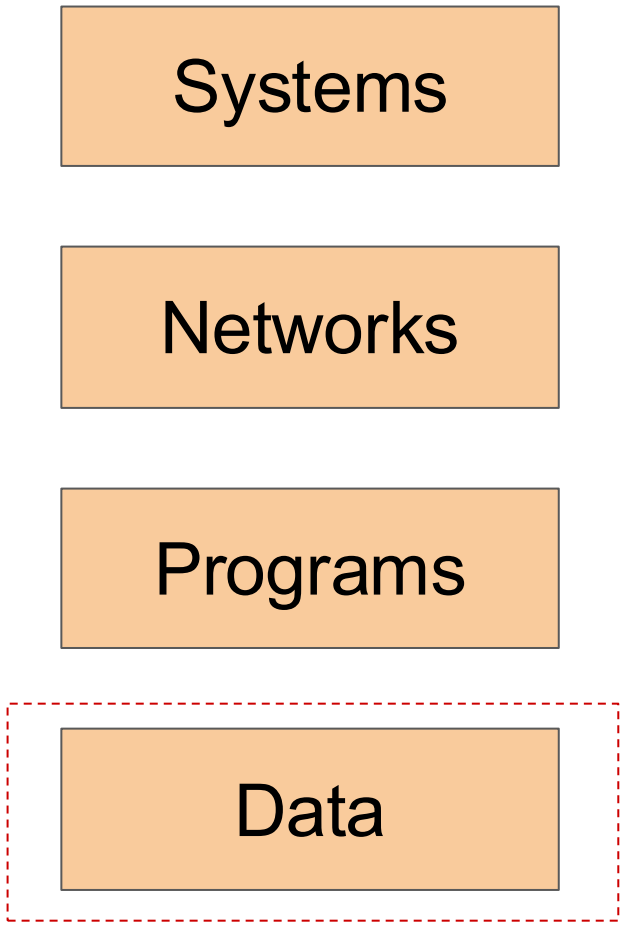
Explicit ethical agents: Representing ethics explicitly.

AI for Privacy:
A Multiagent Perspective

Cybersecurity



Privacy

"the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."
(...)

- Alan Westin

Preserving Privacy in an Online World

How to represent the actual privacy preferences of users?

How to elicit the privacy preferences from users?

How to advise the users to take actions to preserve their privacy?

How to agree on how a co-owned content will be shared?

How to explain privacy decisions?

Real Life Scenarios*

Our online survey with 330 participants shows that more than 90% of privacy violations occur through inference.

** sources: http://www.oddee.com/item 96937.aspx; http://abcnews.go.com/Technology/celebrity-stalking-online-photos-give-location/story?id=11162352*

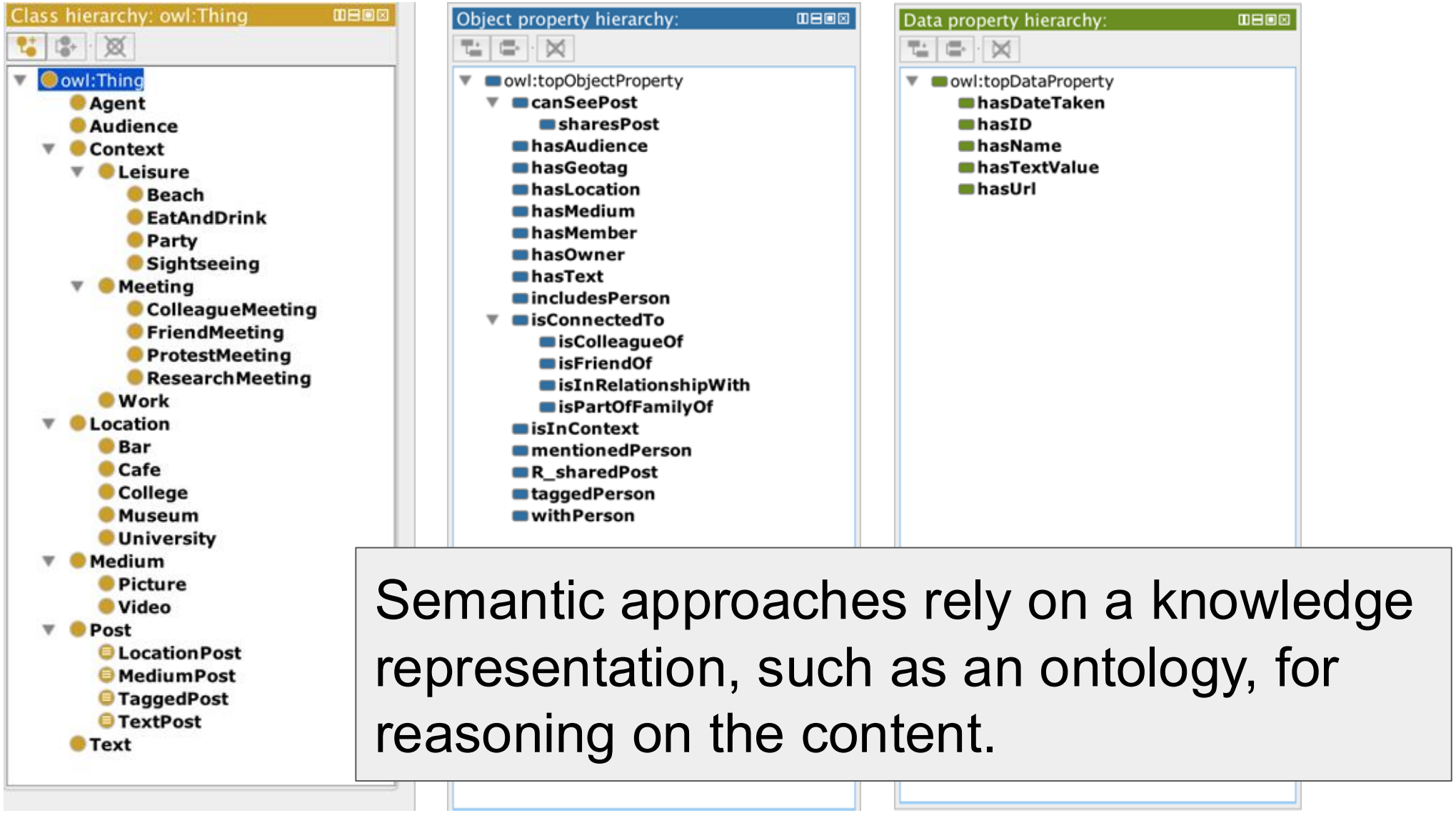
Understanding privacy violations

Privacy Concerns of Dennis

Dennis wants his friends to see his pictures but not his location.

	No inference	**Inference**
User	(i) Dennis checks in at a restaurant.	(iii) Dennis shares a picture without declaring his location. It turns out that his picture is geo-tagged.
Others	(ii) Charlie shares a picture with everyone. He tags Dennis in it as well.	(iv) Charlie checks in at a restaurant. At the same time, Dennis shares a picture of Charlie.

Content Ontology

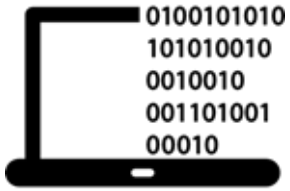


Semantic approaches rely on a knowledge representation, such as an ontology, for reasoning on the content.

Privacy Preferences

P_{E_2}: *hasMedium*(?pr, ?m), *taggedPerson*(?m, :eve),
isInContext(?pr, ?ctx), Work(?ctx) → *rejects*(:eve, ?pr)
[Eve rejects posts that are in work context.]

We can build software agents that can reason on users' privacy preferences.



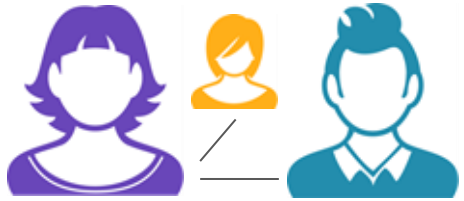
Detection of Privacy Violations: PriGuard

- We represent the social network as an *agent-based online social network* (ABSN).
- Agents know the privacy preferences of their users.
- We develop a sound and complete algorithm to detect privacy violations.
- We show the scalability of the approach on real-life social networks.

PriGuard can detect privacy violations and notify the users to take actions.

Nadin Kökciyan and Pinar Yolum. 2016. "PriGuard: A Semantic Approach to Detect Privacy Violations in Online Social Networks," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 10, pp. 2724-2737

Prevention of Privacy Violations: PriArg

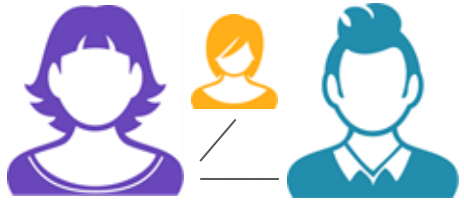


- Agents discuss on a post *before* it is shared.
- We develop a framework that enables agents to carry out a dialogue with other agents.
- We adapt **computational argumentation** to enable privacy decision-making.

Argumentation serves as a useful technique to mimic how humans deal with privacy disputes.

Nadin Kökciyan, Nefise Yaglikci, and Pinar Yolum. 2017. An Argumentation Approach for Resolving Privacy Disputes in Online Social Networks. *ACM Transactions on Internet Technology* 17, 3, Article 27 (June 2017), 22 pages.

Prevention of Privacy Violations: PriNego

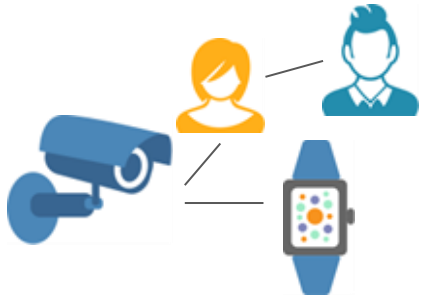


- PriNego is a negotiation-based approach where agents negotiate with each other on their privacy preferences.
- Agents use different negotiation strategies to preserve their users' privacy.
- It exploits reciprocity as a heuristic (e.g., this time you help me, next time I help you).

Agreement can be established over multiple posts.

Dilara Kekulluoglu, Nadin Kökciyan, and Pinar Yolum. 2018. Preserving Privacy as Social Responsibility in Online Social Networks. *ACM Transactions on Internet Technology* 18, 4, Article 42 (April 2018), 22 pages.

Proactive Agents: an IoT example



- Each IoT entity follows **contextual norms** to calculate the appropriateness of sharing information.
- **Computational argumentation** enables the agent to reason on its knowledge and belief bases under **uncertainty**.
- To make inference based on others' information, a **trust model** needs to be in place.

Agents can choose to violate privacy for a better outcome!

Nadin Kökciyan and Pinar Yolum, 2017, August. Context-Based Reasoning on Privacy in Internet of Things. In *IJCAI* (pp. 4738-4744).

Preserving Privacy in an Online World

How to represent the actual privacy preferences of users?

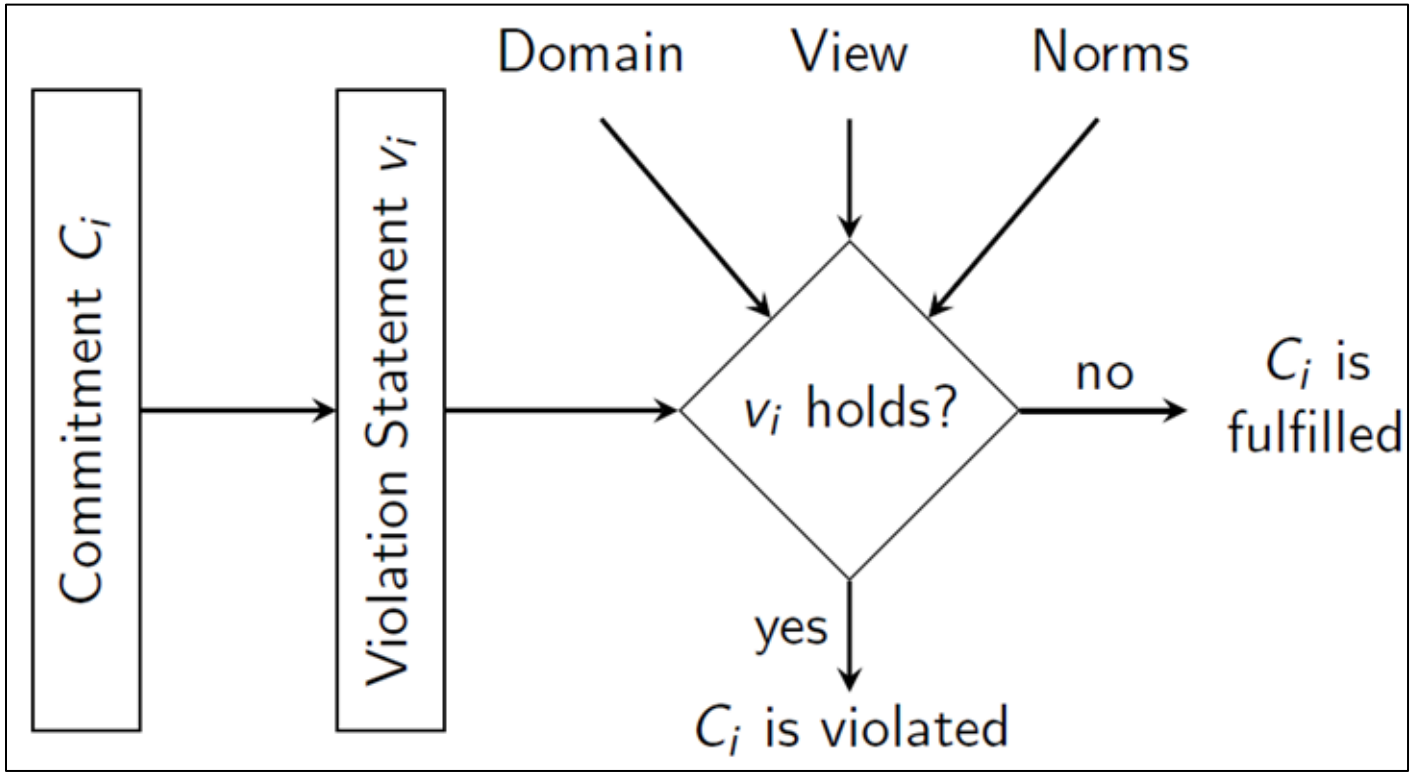
How to elicit the privacy preferences from users?

How to advise the users to take actions to preserve their privacy?

How to agree on how a co-owned content will be shared?

How to explain privacy decisions?

PriGuard: Detection of Privacy Violations



Nadin Kökciyan and Pinar Yolum. 2016. "PriGuard: A Semantic Approach to Detect Privacy Violations in Online Social Networks," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 10, pp. 2724-2737

An Example

Dennis wants his friends to see his pictures but not his location. He posts a picture without declaring his location. However, it turns out that his picture is geotagged.

↓

C_1(:osn, :dennis, isFriendOf(:dennis, X), isAbout(P, :dennis), LocationPost(P), not(canSeePost(X,P)))

↓

V_1 - :osn, :dennis, isFriendOf(:dennis, X), isAbout(P, :dennis), LocationPost(P), canSeePost(X,P))

↓

```
SELECT ?x ?p WHERE {
?x osn:isFriendOf osn:dennis .
?p osn:isAbout osn:dennis .
?p rdf:type osn:LocationPost .
FILTER EXISTS (?x osn:canSeePost ?p) }
```

The Social Network Domain

$\texttt{Agent}, \texttt{Post}, \texttt{Audience}, \texttt{Context}, \texttt{Content} \sqsubseteq \top$	$\texttt{Leisure}, \texttt{Meeting}, \texttt{Work} \sqsubseteq \texttt{Context}$
$\texttt{Beach}, \texttt{EatAndDrink}, \texttt{Party}, \texttt{Sightseeing} \sqsubseteq \texttt{Leisure}$	$\texttt{Bar}, \texttt{Cafe}, \texttt{College}, \texttt{Museum}, \texttt{University} \sqsubseteq \texttt{Location}$
$\texttt{Picture}, \texttt{Video} \sqsubseteq \texttt{Medium}$	$\texttt{Medium}, \texttt{Text}, \texttt{Location} \sqsubseteq \texttt{Content}$
$\texttt{Post} \sqcap \exists \mathit{sharesPost}^{-}.\texttt{Agent} \equiv \exists \mathit{R\_sharedPost}.\mathit{Self}$	$\texttt{LocationPost} \equiv \exists \mathit{R\_locationPost}.\mathit{Self}$
$\texttt{LocationPost} \equiv \texttt{Post} \sqcap \exists \mathit{hasLocation}.\texttt{Location}$	$\texttt{MediumPost} \equiv \texttt{Post} \sqcap \exists \mathit{hasMedium}.\texttt{Medium}$
$\texttt{TaggedPost} \equiv \texttt{Post} \sqcap \exists \mathit{isAbout}.\texttt{Agent}$	$\texttt{TextPost} \equiv \texttt{Post} \sqcap \exists \mathit{hasText}.\texttt{Text}$

Norms

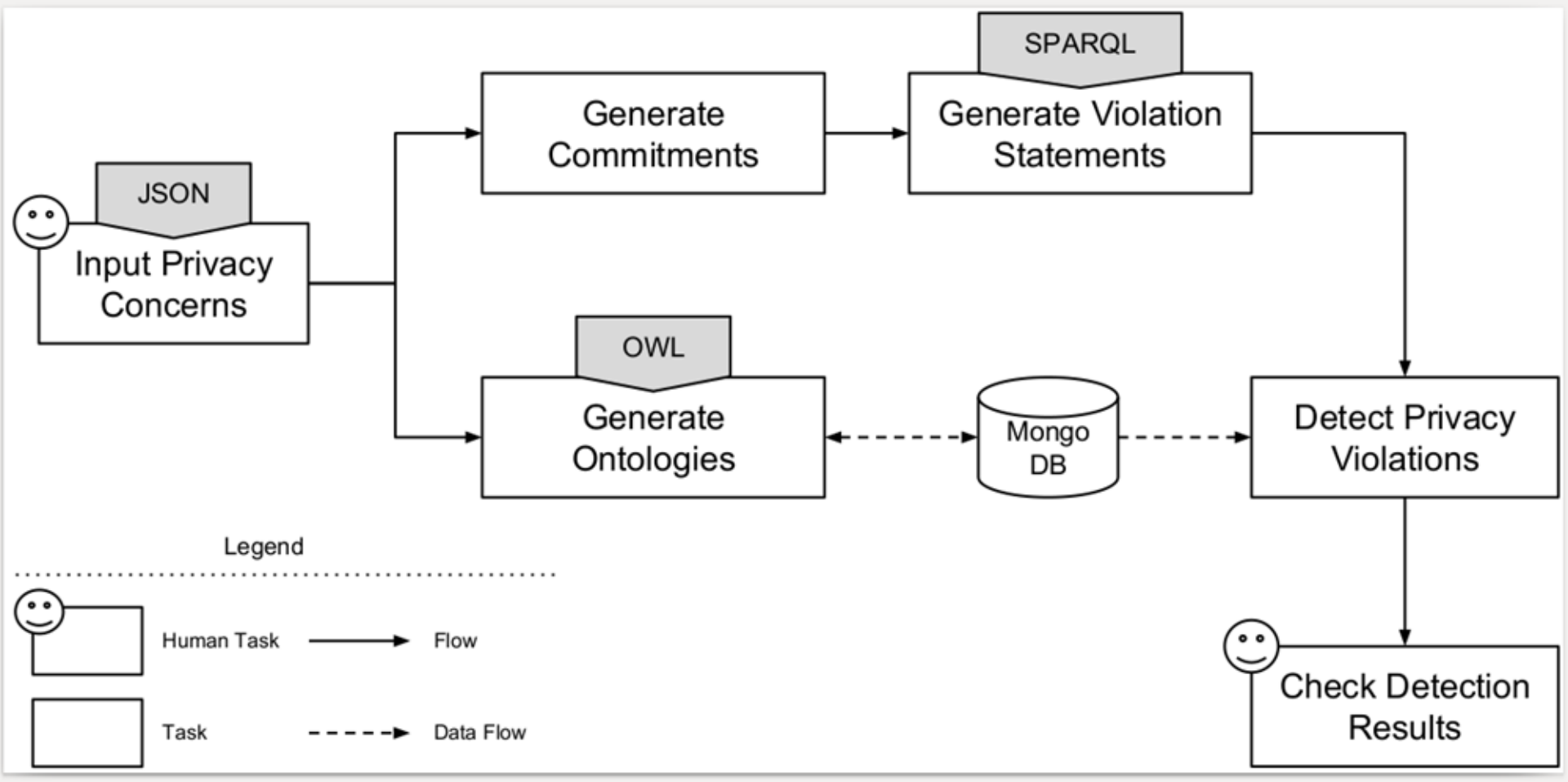
N_1: $sharesPost(X,P) \rightarrow canSeePost(X,P)$
[Agent can see the posts that it shares.]

N_2: $sharesPost(X,P) \wedge hasAudience(P,A) \wedge hasMember(A,M) \rightarrow canSeePost(M,P)$
[Audience of a post can see the post.]

N_3: $hasMedium(P,M) \wedge taggedPerson(M,X) \rightarrow isAbout(P,X)$
[Post is about agents tagged in a medium.]

N_4: $\texttt{Post}(P) \wedge hasMedium(P,M) \wedge hasGeotag(M,T) \rightarrow \texttt{LocationPost}(P)$
[Geotagged medium gives away the location.]

A Facebook Application: PriGuardTool



Kökciyan N., Yolum P. (2016) PriGuardTool: A Web-Based Tool to Detect Privacy Violations Semantically. In: Baldoni M., Müller J., Nunes I., Zalila-Wenkstern R. (eds) Engineering Multi-Agent Systems. EMAS 2016. Lecture Notes in Computer Science, vol 10093. Springer, Cham