

The Role of Online Forums in Developer Understanding of Privacy Law – A Reddit Case Study

Sara Haghighi¹, Clark LaChance¹, Ali Pourghasemi Fatideh¹, Travis Breaux², Sepideh Ghanavati¹
University of Maine¹, Carnegie Mellon University²



Introduction

- Software practitioners use online forums to navigate complex and ambiguous legal privacy requirements, such as GDPR¹.
- Little is known about their professional backgrounds, challenges faced, how they assess the credibility of advice, or how they resolve ambiguities in posts.
- We conducted a survey of **223 Reddit users** from regulatory-focused subreddits, complemented by a qualitative analysis of **2,248 posts** and their responses.
- Despite holding privacy certifications, most participants frequently use forums to seek legal advice.
- Key challenges reported or identified include implementing a data protection impact assessment, reporting a data breach, and obtaining cookie consent.
- Reddit users often assess credibility by reviewing respondents’ post history, verifying sources cited, trusting advice from recognized experts, and following up for clarity before responding.

Objectives

- What motivates users to **engage in privacy and legal discussions** on regulatory-focused subreddits, and does this vary based on professional and organizational background?
- What are the **most common challenges** users face when seeking compliance or privacy-related advice, and how do they vary by professional and organizational factors?
- How do users **assess the credibility of information** shared, **resolve ambiguities** in posts, and how do these assessments affect their decision-making processes? Does perceived credibility vary based on professional background?
- How do the challenges reported by Reddit **users** compare with those identified in the **posts**?

Methods

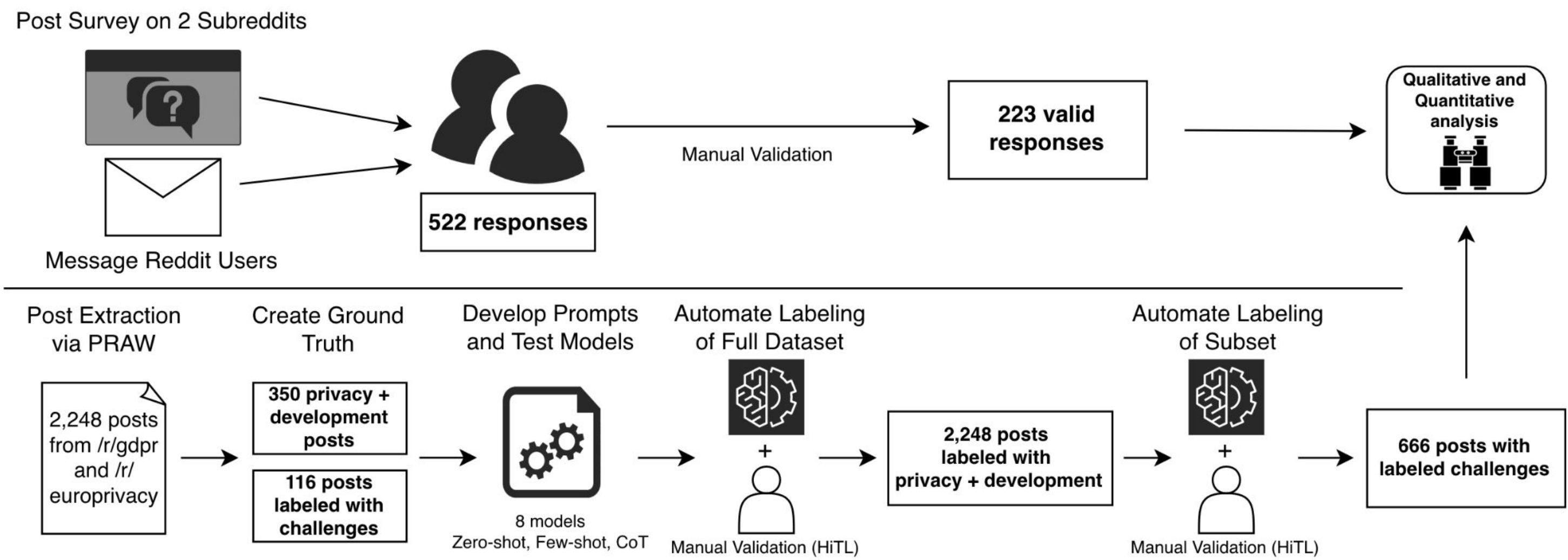
Survey

- We designed the survey through discussions with legal experts and conducted a pilot study.
- We recruited users from regulatory-focused subreddits.
- 522 responses were collected. We filtered them to ensure high quality, yielding **223 valid responses**.

Reddit Posts

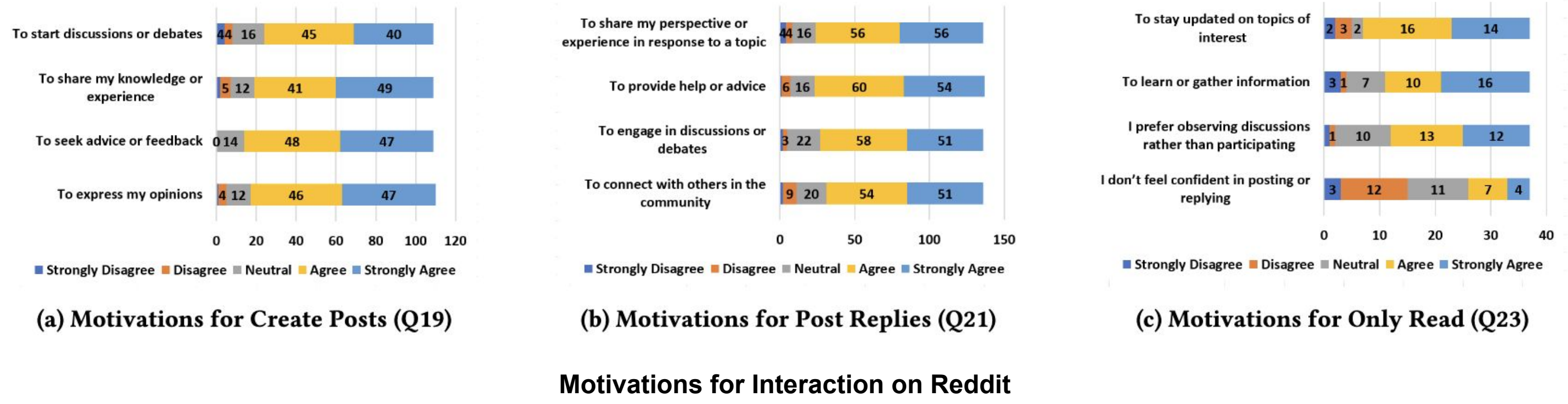
- Using PRAW, we conducted a keyword search to extract 2,248 posts from /r/GDPR and /r/europrivacy.
- We created a ground truth dataset, labeling each post as privacy-related, and if so, what GDPR-related challenges are present.
- We constructed CoT-based prompts and use GPT5 to label the full dataset.
- We used a human-in-the-loop approach² to verify the model annotations.
- This process yielded **666 posts** annotated with the GDPR challenges present.

Methods



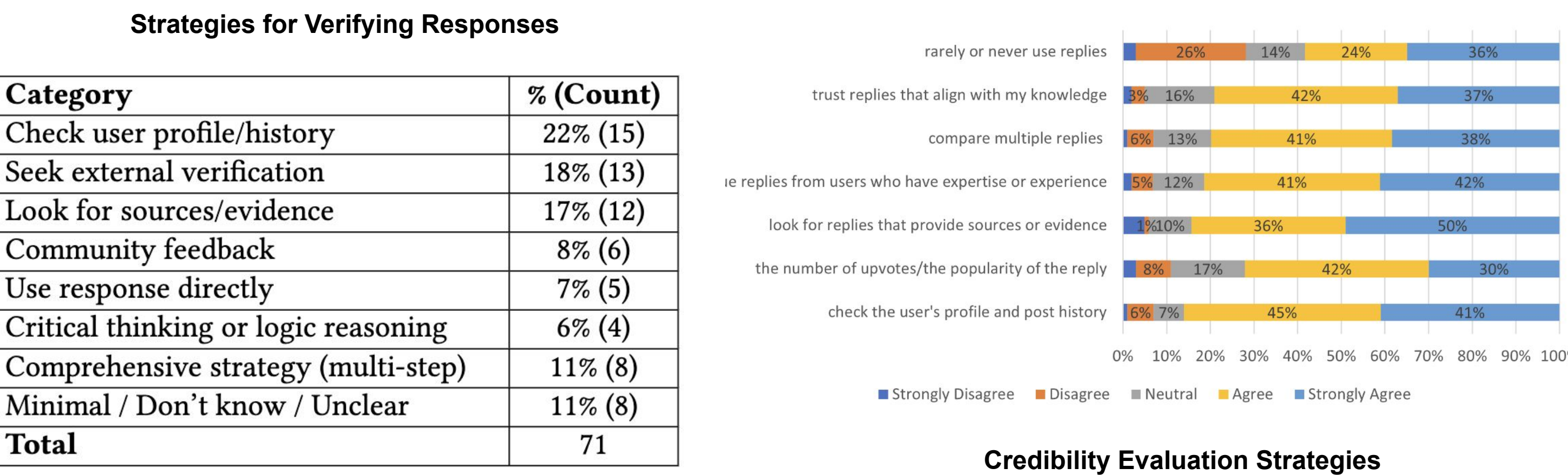
Results

- Users ask a range of questions about GDPR, including those related to DPAs, obtaining consent, DPIAs, DSARs, and RoPA.
- One user asked “*if a DPA is necessary if their privacy policy and terms of service documents contain the same information.*”
- Another user sought advice on “*access to sensitive data with remote workers in other countries.*”



GDPR Challenges in Annotated Reddit Posts	
* Indicates GDPR challenges also identified in the survey.	
Category	% (Count)
Cookie Consent	20% (132)
General Consent	18% (119)
Data Processing Agreements (DPAs)	12% (81)
Data Transfer	12% (77)
Other	9% (62)
GDPR Definitions	8% (55)
Data Protection Impact Assessment (DPIA)*	6% (39)
Data Subject Access Requests (DSARs)*	6% (38)
Generic Compliance	5% (32)
Records of Processing Activities (RoPA)*	4% (24)
Data Breach Notification*	1% (10)
Total GDPR Challenges	100% (669)

Distribution of GDPR Challenges in Survey	
Category	% (Count)
Data Protection Impact Assessment (DPIA)	55% (121)
Data Breach Notification	49% (107)
Data Subject Access Requests (DSARs)	39% (85)
Records of Processing Activities (RoPA)	29% (64)
Other	1% (3)



Discussion

Findings

- Regulatory-focused subreddits support both advice-seeking and knowledge-sharing around compliance.
- Privacy-related experience, certifications, and organizational context may shape how users participate in discussions.
- Challenges vary by company type/size, and holding a privacy-related certification. Most companies rely on external legal help to fill gaps in their internal privacy expertise.
- Most users assess credibility using evidence-based methods like verifying external sources, checking profile histories, or upvotes, and seek further clarification for ambiguous posts.
- Selecting one strategy over another is explained by some professional backgrounds, but privacy experience does not affect such decisions.
- Our analyses show there is high motivation to provide advice (83%) and to share developer perspectives and experiences (82%).

Implications for Educators and Companies

- There is a need for companies and educators to create training materials on topics such as consent, PII identification, data processing agreements, data transfers, and broader regulatory compliance.
- Better tooling, like open-source SDKs with privacy-related components or privacy notice generators that draw from source code, could help mitigate common challenges.

Future Research Directions

- Future work could explore using human-AI collaboration on implementing PbD practices.
- More research could examine how forum users surface assumptions and facts in posts and resolve misunderstandings/ambiguities during discussions.

Acknowledgements

We would like to acknowledge the support provided by NSF CAREER #2238047 and NSF Award #2217572.

References

- European Union. The EU General Data Protection Regulation (GDPR). <http://www.eugdpr.org/>. Accessed February 10, 2025.
- Nicholas Pangakis and Samuel Wolken. 2024. Keeping Humans in the Loop: Human-Centered Automated Annotation with Generative AI. *arXiv preprint arXiv:2409.09467* (2024).