Group Name - All4s

Group Members

Kasim Taylor

Giselle Shah

Michelle Khadan

Exploring the Association between Moral Foundations and Judgements of AI Behaviour

Published in the events of the CHI conference on

May 11-16, 2024

Done By Kasim Taylor

Meet the authors

Eduardo Velloso
University professor with a PhD in computer science. His background is design, engineering and psychology

Frank Vertere
An alumni of the University of Melbourne who obtained his PhD. his expertise in the HCI allowed him to receive 7 million dollars in funding for his research

Joe Brailsford
A PhD student who work in the Human-Computing lab. His research is on factors that can help shape the design of responsible product policy guidelines

What's it about??

Objective:

- Investigating how individual moral values affect perception of AI behaviour in morally contentious situation.

Methodology:

- Applied Moral Foundation Theory(MFT)
- Conducted the study with 240 participants.
- Participants judge the AI's behaviour in six different scenarios.
- Bayesian modeling and thematic analysis used to assess moral judgments..

What's it about(cont.)

Key findings:

- Mild association between individual MFT scores and judgements of AI.
- Technical understanding of AI, rather than moral values, influenced participants' judgements.
- Participants who saw AI as more human-like judges it more harshly.

Implications:

- Important for designing AI systems that align with user perceptions and ethics in Human-Computer Interaction.

Methodology

Research Methods:

- Applied the **Moral Foundation Theory(MFT)** to understand moral judgements of AI behaviour.
- Used a combination of **quantitative**(Bayesian modeling) and **qualitative**(thematic analysis) approaches.

Experimental Design:

- 240 participants were recruited via the Prolific platform.
- They were shown **6 AI-related scenarios** covering different moral tensions.
- Each scenario was used to trigger responses related to **6 moral foundations.**
- Participants completed the **Moral Foundations Questionnaire(MFQ)** to assess their sensitivity to each foundation.

Methodology(cont.)

Data Collection

- Quantitative: participants rated the AI behaviour using a 100-point scale to judge how morally right or wrong the behaviour was.
- Qualitative: participants provided **open-ended responses** explaining their judgement.
- Bayesian statistical analysis used to model participants' **moral judgements** and how they are related to their MFQ scores.

Usability/Analysis:

- **Kendall's coefficient of concordance** was used to measure agreement among participants about the relevant moral foundations.
- **Thematic analysis** of qualitative responses to identify recurring themes related to AI perception and moral foundation.

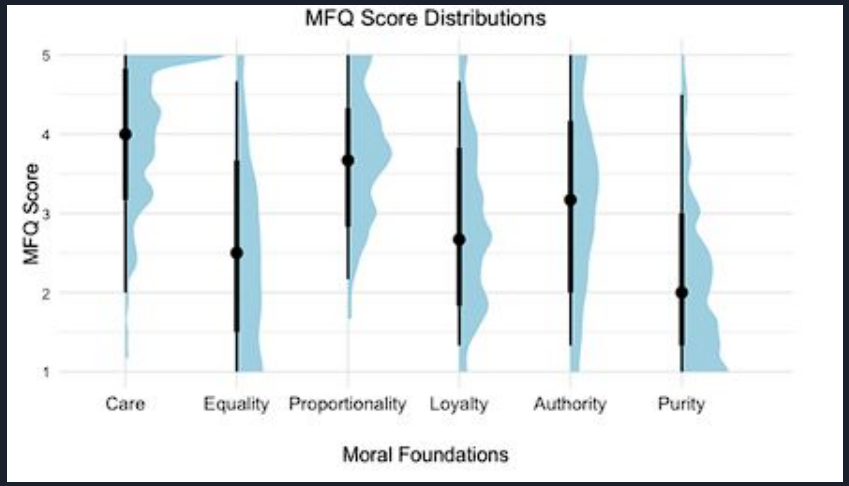
Results

Key findings

- **Weak agreement** among participants on which moral foundations were most relevant to AI behaviour.
- **Care** foundation was the only consistent predictor of how participants judged AI behaviour as right or wrong
- **Technical understanding** of AI influenced participants' judgements more than their moral foundation scores.
- Participants with lower technical understanding tended to **anthropomorphize** AI systems, leading to harsher judgement.

Results(cont.)

Figure 2 shows the distribution of the Moral Foundations Questionnaires(MFQ) scores across participants, showing their sensitivity to various moral foundations like care, authority, loyalty, etc.



Results(cont.)

Table 1 shows the **agreement levels** from Kendall's W across different scenarios. Healthcare being W= 0.31, education being W=0.32 and criminal justice being W=0.15, just to name a few. What this means is that there is a slight to fair agreement on the others, with a lower agreement to the criminal justice scenario.

Scenario	*W*	Interpretation
Healthcare	0.31	Fair agreement
Education	0.32	Fair agreement
Employment	0.34	Fair agreement
Social Services	0.27	Fair agreement
Financial Services	0.32	Fair agreement
Criminal Justice	0.15	Slight agreement

Results(cont.)

Table 2 shows the relationship between individual MFQ scores and the perceived relevance of moral foundations in each scenarios. A higher MFQ score leads to a higher likelihood of identifying that foundation as relevant

Effect	Est. (SD)	89% CI
$MFQ_{Auth} \rightarrow Scenario_{Auth}$	.06 (.03)	[.02, .11]
$MFQ_{Pure} \rightarrow Scenario_{Pure}$	.24 (.03)	[.19, .30]
$MFQ_{Prop} \rightarrow Scenario_{Prop}$	.11 (.04)	[.05, .17]
$MFQ_{Care} \rightarrow Scenario_{Care}$	.07 (.03)	[.02, .12]
$MFQ_{Loyal} \rightarrow Scenario_{Loyal}$	.15 (.03)	[.10, .20]
$MFQ_{Equal} \rightarrow Scenario_{Equal}$	.00 (.03)	[-.05, .04]

Results(cont.)

Table 3 shows that **care** is the most significant foundation predicting the moral judgements whereas the other had less of a impacted.

Parameter	Estimate	SD	89% CI
Ordinal regression cut-points			
$\tau[1]$	-1.37	0.54	[-2.23, -0.50]
$\tau[2]$	-0.50	0.53	[-1.35, 0.35]
$\tau[3]$	0.81	0.53	[-0.04, 1.66]
$\tau[4]$	2.23	0.54	[1.37, 3.08]
Individual MFQ scores			
Care	0.24	0.08	[0.12, 0.36]
Equality	0.09	0.07	[-0.02, 0.21]
Proportionality	0.00	0.10	[-0.16, 0.16]
Loyalty	-0.15	0.10	[-0.31, 0.02]
Authority	0.16	0.10	[-0.01, 0.32]
Purity	-0.01	0.08	[-0.14, 0.12]
Scenario MFT ratings			
Care	-0.04	0.03	[-0.08, 0.01]
Equality	-0.00	0.03	[-0.08, 0.04]
Proportionality	-0.05	0.02	[-0.09, -0.01]
Loyalty	-0.08	0.03	[-0.13, -0.03]
Authority	-0.04	0.03	[-0.09, 0.00]
Purity	-0.03	0.03	[-0.08, 0.02]

But what does all of this mean??

Implications in the field:

- **Technical understanding** plays a large role in how users judge AI behaviour. This suggests that designers should put a importance on transparency about the system to improve perception.
- **Anthropomorphism** can impact moral judgements. AI system should be designed in a way that it is clear about it's non-human attributes to prevent misjudgement.

But what does this mean??(cont.)

Limitation and shortcomings:

- **Limited scenarios scope:** although the scenarios are realistic, it doesn't cover the wide range of situations that can appear.
- **Geographical limitation:** participants were mostly from the USA and UK so it doesn't take into account the various cultures.
- **Controlled environment:** since these scenarios happen in a controlled place, it limits how users may make moral decision in the real world therefore making the real-life applicability of the findings small.

Conclusion

Main takeaways:

- Moral judgements of Ai behaviour are influenced more by **technical understanding** than an individual's moral values.
- The **care** foundation was the predictor of participants' moral judgements.
- People with a **lower technical understanding** of AI tends to anthropomorphize them and judge them harshly.
- The study shows that people's relation to AI is evolving, and with varying degrees of moral responsibility assigned to AI system.

Conclusion

Relevance and impact:

- AI needs to be presented in a way to reduce the anthropomorphism and clarify their computational nature.
- Highlights the need for more nuanced AI ethics guidelines that consider human diversity in moral judgements.
- Influence future HCI and AI ethics research, focusing on the role of technical understanding in shaping human-AI interaction.

References

Ali, S. J., Christin, A., Smart, A., & Riitta Katila. (2023). Walking the Walk of AI Ethics: Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs. *ArXiv (Cornell University)*. https://doi.org/10.1145/3593013.3593990

APA PsycNet. (2024). Apa.org. https://psycnet.apa.org/record/2020-59494-001

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI We trust? Perceptions about Automated decision-making by Artificial Intelligence. *AI & SOCIETY*, *35*(3). https://doi.org/10.1007/s00146-019-00931-w

Bhatt, U., Andrus, M., Weller, A., & Xiang, A. (2020). Machine Learning Explainability for External Stakeholders. *ArXiv:2007.05408 [Cs]*. https://arxiv.org/abs/2007.05408

“I Don’t Even Remember What I Read”: How Design Influences Dissociation on Social Media

BACKGROUND

The article was published during the CHI '22: CHI Conference on Human Factors in Computing Systems which took place in New Orleans LA USA from 29 April 2022- 5 May 2022.

The article examines how social media design affects users' cognitive absorption.

A custom Twitter client, Chirp, led to users experiencing dissociation, with mixed feelings about its value. Social media designs exploit users' tendencies for normative dissociation. However, incorporating internal and external supports in design can reduce the chances of such dissociative experiences.

Done By Giselle Shah

AUTHORS

Amanda Baughan- Developing design features for a messaging platform to enhance connection and understanding in conflict and emotional conversations.

Mingrui “Ray” Zhang- Researched and improved technology to enhance and make human communication more accessible.

Raveena Rao- Worked on Tutela app to enhance safety for women using Uber or walking home.

Kai Lukoff- research investigates how digital interfaces impact user experience and sense of agency.

Anastasia Schaadhardt- Developed creativity tools for the visually impaired and conducted accessibility research with Microsoft Research's Enable team.

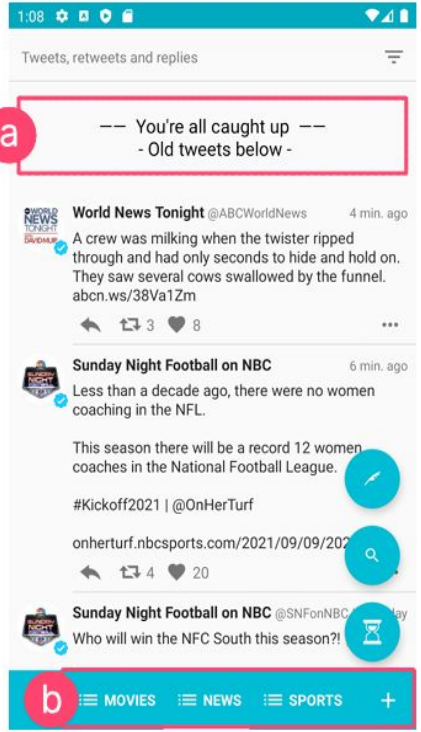
Lisa Butler- has published widely on various aspects of trauma, resilience, growth, and dissociation.

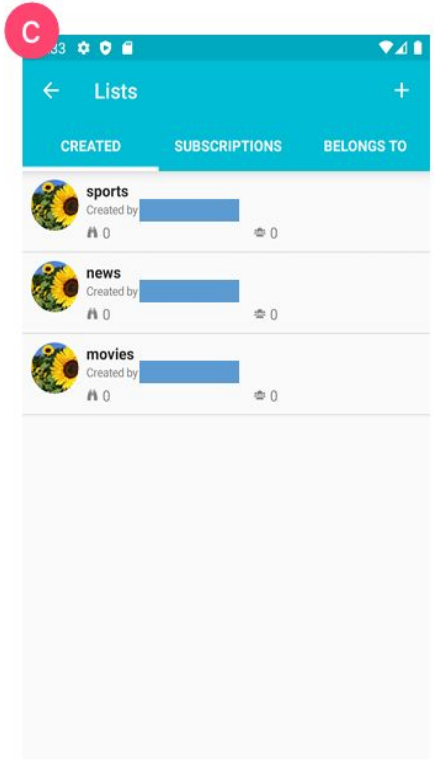
Alexis Hiniker- investigates the ways in which everyday technologies make life worse for their users.

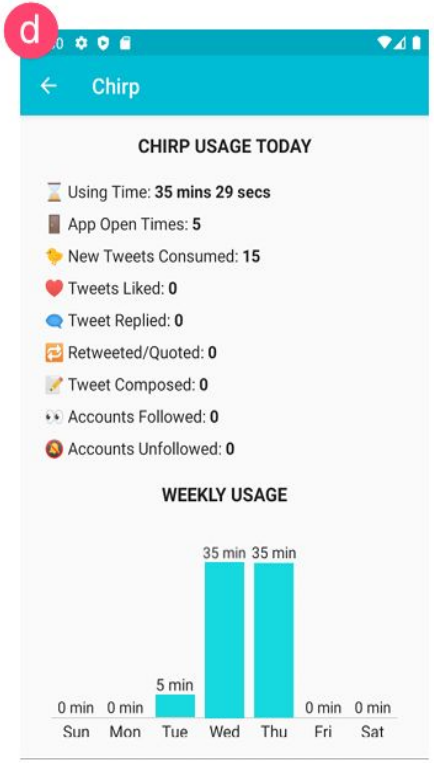
METHODOLOGY

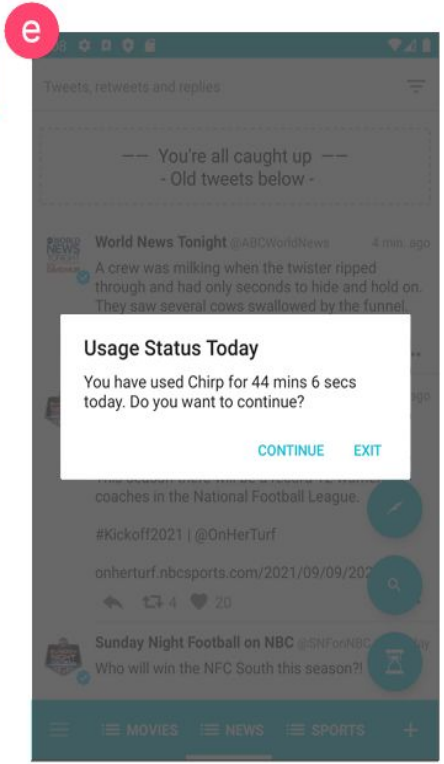
- Four versions of a custom Twitter client were created, called Chirp.
- Internal interventions were used to change how users consume tweets, and external interventions gave users tools to monitor their use of the app.
- At the conclusion of deployment, interviews were conducted.
- User was prompted to respond to the statement “I am currently using Chirp without really paying attention to what I am doing.” using a scale of 1 (Strongly Disagree) through 5 (Strongly Agree).

ILLUSTRATION

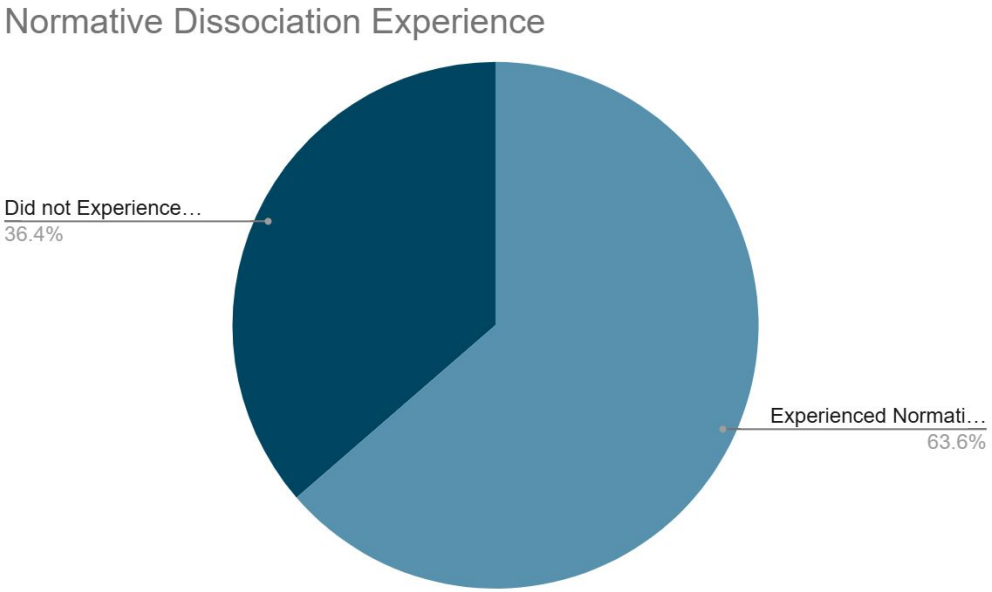








RESULTS



- 7 of the 11 interviewees described experiencing normative dissociation when using social media.

- 18 of the 43 deployment participants responded at least once to the questionnaire prompt by **agreeing** or **strongly agreeing** with the statement that they were using Chirp without paying attention to what they were doing, for a total of 58 instances of normative dissociation.

RESULTS

- Participants explained that Twitter's evolving design and infinite feed of posts aid in getting them hooked on using the app.
- Two users did not think it fit the "mindless" label, some saw it as inevitable, and others felt shame and anger after realizing how much time they had "wasted."
- Users reported that the internal and external interventions helped them to reduce their mindless consumption of Chirp.

DISCUSSION

- The findings show that social media use often leads to dissociation, whether through **deep absorption** or **mindless scrolling**. The article suggests that this should not be considered addiction but as "normative dissociation," a natural process of seeking escape.
- It argues that social media design fosters prolonged dissociation, causing frustration and dissatisfaction. The article concludes by arguing that it is possible for users to have healthy and satisfying relationships with social media, even while dissociating, if platforms provide a pathway to disengagement.

CONCLUSION

People described their social media use in ways that fit the normative dissociation model: either becoming deeply absorbed in their content consumption, or mindlessly scrolling while absorbed in other thoughts.

However, the design can reduce and disrupt normative dissociation, or encourage positive disengagement, maximizing the benefits of normative dissociation on social media and prompt self-awareness.

REFERENCES

1. Amanda Baughan, A. x, Zhang , M. “Ray”, Rao , R., Lukoff, K., Schaadhardt , A., Butler, L., & Hiniker, A. (2022). *“I Don’t Even Remember What I Read”: How Design Influences Dissociation on Social Media* . *18*, 1–13. https://doi.org/10.1145/3491102.3501899

2. Baughan, A. (2024). *Amanda Baughan*. Amandabaughan.com. https://www.amandabaughan.com/

3. Lukoff, K. (2023). *Kai Lukoff – HCI researcher & teacher*. Kailukoff.com. https://kailukoff.com/

4. Rao, R. (2021). *Raveena Rao’s Portfolio*. Raveenarao.com. https://www.raveenarao.com/

5.Zhang, M. (2024). *Mingrui (Ray) Zhang · DrustZ*. Drustz.com. https://drustz.com/

6.Schaadhardt, A. (2021). *Anastasia’s Homepage*. Anastasia’s Homepage. https://www.anastasiaschaadhardt.com/

7.Lisa Butler. (2023, May 4). Buffalo.edu. https://socialwork.buffalo.edu/faculty-research/full-time-faculty/ldbutler.html

8. iSchool Directory | Information School | University of Washington. (2017). Uw.edu; UW iSchool. https://ischool.uw.edu/people/faculty/profile/alexisr

"I Deleted It After the Overturn of Roe v. Wade":

Understanding Women's Privacy Concerns Toward Period-Tracking Apps in the Post Roe v. Wade Era

Done By Michelle Khadan

Background

Authored by Jiaxun Cao, Hiba Laabadli, Chase Mathis, Rebecca Stern, and Pardis Emami-Naeini, all of whom are affiliated with Duke University in the United States

This paper was published in the CHI '24: CHI Conference on Human Factors in Computing Systems in May 2024 in Honolulu, Hawaii which ran from the 11th to the 16th

Abstract

The overturn of Roe v. Wade has taken away the constitutional right to abortion. Prior work shows that period-tracking apps' data practices can be used to detect pregnancy and abortion, hence putting women at risk of being prosecuted. It is unclear how much women know about the privacy practices of such apps and how concerned they are after the overturn. Such knowledge is critical to designing effective strategies for stakeholders to enhance women's reproductive privacy. We conducted an online 183-participant vignette survey with US women from states with diverse policies on abortion. Participants were significantly concerned about the privacy practices of the period-tracking apps, such as data access by law enforcement and third parties. However, participants felt uninformed and powerless about risk mitigation practices. We provide several recommendations to enhance women's privacy awareness toward their period-tracking practices.

Methodology

The researchers conducted a vignette-based study to collect both quantitative and qualitative data. This method has been commonly used in other privacy-related research to understand how people make decisions about privacy in different contexts.

Rather than only using a fixed definition of privacy, the use of the vignette method allows the researchers to explore how various factors influence people's decisions about privacy

Methodology

Using previous research on privacy concerns, the researchers decided to test these four factors;

- **Collected data:** The type of data the period-tracking app collects
- **Data storage:** Where the data is stored.
- **Data sharing:** Who the data is shared with
- **User control:** The level of control users have over their data

In the survey, participants were presented with short hypothetical scenarios (vignettes). For each vignette, these factors were combined in different ways, and each participant was presented with four randomly selected scenarios out of a total of 75 possible combinations.

Methodology

Using the online survey site Prolific, the researchers recruited 200 participants that met the following requirements;

- Must be female (self-identified as cisgender or transgender).
- Must be at least 18 years old.
- Must live in the United States.
- Must have an approval rating of over 95% on Prolific.

In order to examine privacy concerns under different legal contexts, the researchers split the participants evenly between states that allowed abortion and states where it was banned. A total of 183 valid responses were able to be collected. Each participant received 4 USD for completing the survey, which took an average of 14 minutes to complete.

Methodology

To summarize the contents of the survey, these were the sections included in it.

- Scenario Presentation
- Follow-up Questions
- Attention-Check Questions
- Period-Tracking App Usage
- Roe v. Wade Impact
- Demographics

Results

Data Sharing

Law enforcement access to period-tracking data was rated as the most concerning, with participants feeling that such sensitive information shouldn't be shared with authorities under any circumstances. Many participants found it unnecessary for law enforcement to access such personal data.

Data Collected

While menstrual cycle data was viewed as necessary for the app's functionality, the collection of location data raised significant concerns. Many respondents saw location tracking as irrelevant to period tracking and were uncomfortable with this data being stored or shared.

Results

User Control and Data Storage

The option to delete data significantly reduced participants' worries. Participants expressed that when they had the ability to delete or control their data, they felt more secure in using the app.

Data storage location as of lesser concern compared to data sharing and user control, though participants still preferred local device storage over cloud storage due to potential security risks.

Results

Impact of Roe v. Wade

Interestingly, **only 8% of participants** explicitly linked their privacy concerns to the overturn of Roe v. Wade, despite the centrality of the event to the study. However, many did mention <u>increased anxiety about the legal implications</u> of using period-tracking apps post-Roe v. Wade.

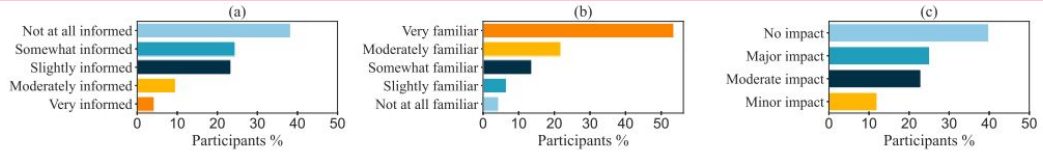


Figure 4: Participants' (a) knowledge level of the privacy practices of their PTA, (b) familiarity with the overturn of Roe v. Wade, and (c) perceived impact on concerns toward the privacy of their PTA post the overturn of Roe v. Wade.

Results

Privacy Mitigation Practices

A majority of the participants took no action to mitigate their privacy concerns, citing a lack of knowledge about how to protect their data. **Only 9% of participants** actively sought to manage their privacy by either deleting apps or reading privacy policies in detail.

Discussion

The research highlights the privacy risks associated with period-tracking apps, especially in the context of the Roe v. Wade overturn. These findings are significant for the HCI field as the emphasize the need for stronger privacy protections in apps. Some recommendations from the researchers include;

- improving data transparency
- offering users granular control over their data (such as deletion options)
- limiting the collection of irrelevant data.

The study also shows how legal factors, particularly in abortion-restrictive states, shape privacy concerns, especially regarding data sharing with law enforcement. Though the study has limitations, mainly its U.S.-centric approach, it still has a broader impact of stressing the importance of privacy-aware design in health technologies.

Conclusion

In summary, this research paper brings attention to the fact that period-tracking apps are often not transparent with users about how their sensitive data is being collected and utilized, leaving many users unaware of the extent to which their personal information is being exploited particularly in the post-Roe context.

The relevance of this research is its ability to shed light on the shady practices of period-tracking apps and advocate for strategies to raise awareness awareness and encourage women to safeguard their privacy.

These insights are essential for empowering women and to put pressure app developers to implement stronger privacy protections for reproductive justice.

References

Cao, Jiaxun, et al. “I Deleted It Afer the Overturn of Roe v. Wade”: Understanding Women’s Privacy Concerns toward Period-Tracking Apps in the Post Roe v. Wade Era Pardis Emami-Naeini. 2024, www.ftc.gov/system/files/ftc_gov/pdf/10-Laabadli-Understanding-Womens-Privacy-Concerns-Toward-Period-Tracking-Apps-in-the-Post-Roe-v-Wade-Era.pdf, https://dl.acm.org/doi/10.1145/3613904.3642042