Multiverse Privacy Theory for Contextual Risks in Complex User-AI Interactions

Ece Gumusel
Indiana University Bloomington

Abstract

In an era of increasing interaction with artificial intelligence (AI), users face evolving privacy decisions shaped by complex, uncertain factors. This paper introduces Multiverse Privacy Theory, a novel framework in which each privacy decision spawns a parallel universe, representing a distinct potential outcome based on user choices over time. By simulating these universes, this theory provides a foundation for understanding privacy through the lens of contextual integrity, evolving preferences, and probabilistic decision-making. Future work will explore its application using real-world, scenario-based survey data.

1 Introduction

As artificial intelligence (AI) technologies increasingly rely on vast datasets for training and operation, the need to protect user privacy has become more critical. This concern extends beyond just tracking user behavior [11, 23]—it also includes protecting against privacy-targeted malicious attempts and adversarial attacks [19], which are becoming easier to execute [3, 12, 20]. In addition, despite the growing urgency to address these threats, many developers still have a limited understanding of user privacy expectations and its broader implications [2, 13]. They often view privacy narrowly as a matter of regulatory compliance, focused mainly on safeguarding personally identifiable information (PII) [8]. However, it is a much more complex issue—one that also involves respecting user expectations, ensuring transparency, and fostering trust and satisfaction in AI systems. Especially in accounting for evolving privacy preferences highlighted in previous literature, the dynamic nature of security and privacy threats, and the diversity of user demographics—such as minors, older adults, individuals with disabilities, LGBTQ+ individuals, or users experiencing mental health challenges—whose privacy concerns, disclosure patterns, and trust dynamics vary significantly from the general population [5, 9, 10, 16]. Thus, user privacy decisions should not be static or one-size-fits-all. Instead, they should be evaluated within the context of multiple potential infinite number of scenarios, each representing a different privacy decision under various circumstances.

To address these risks, notable privacy–utility trade-off models have been developed, aiming to balance data usefulness with effective privacy protection. For example, differential privacy (DP) ensures the risk of identifying any individual in a dataset remains low by analyzing pairs of neighboring databases differing by one record [7]. Statistical Data Disclosure (SDC) techniques vary in privacy strength depending on database size and influence later data analysis. Slavković and Seeman's Statistical Data Privacy (SDP) framework [22] also offers an analysis of data release mechanisms that sanitize outputs based on confidential data while considering broader statistical disclosure risks by extending SDC and DP.

Contextual Integrity (CI) [18] also provides a critical lens of privacy frameworks by defining privacy as the appropriateness of information flows, emphasizing that privacy norms are context-dependent and shift based on the social and informational settings in which data are used. In other words, privacy is preserved when information flows align with the contextual norms in governing 5-parameters: actors (senders, recipients, and subjects), transmission principles or condition, information type, information norms, and user groups [18]. Recent work has also integrated the governing knowledge commons framework (GKC) with CI—forming the unified GKC-CI model—which examines how information flows are shaped by institutional rules, social roles, and shared resources [21].

However, current privacy models typically operate under static assumptions and struggle to capture the full spectrum of user expectations, CI and/or GKC-CI factors, and system-

Accepted to Societal & User-Centered Privacy in AI (SUPA) at USENIX Symposium on Usable Privacy and Security (SOUPS) 2025.
August 10–12, 2025, Seattle, WA, United States.

specific risks. These models often prioritize measurable outcomes—such as data minimization or risk scores—while overlooking less quantifiable but equally critical aspects like user trust, perceived control, and the dynamic interplay between evolving privacy harms, risks, and threats. As a result, a significant gap persists between theoretical privacy protections and practical, user-centered outcomes in real-world AI systems. To be effective in dynamic, data-intensive environments—where privacy expectations are fluid, security risks evolve rapidly, and user demographics are increasingly diverse—these models require further extension. Despite various efforts to address these challenges [1, 4, 14, 15], their primarily normative focus often falls short in operationalizing privacy within technical systems that demand continuous adaptation, personalization, and real-time decision-making [24]. As such, there is a growing need for integrative approaches that bridge conceptual insights with actionable mechanisms for privacy-aware AI design.

Our Approach. To overcome these limitations, we propose the Multiverse Privacy Theory (MPT) integrates principles from multiverse [6, 17] and CI theories [18] to better handle complex computational systems and diverse user attributes. The idea behind applying multiverse theory to user privacy decision-making within computational systems revolves around the concept of decision-making across multiple potential time-dependent realities or scenarios, each of which reflects different choices and outcomes based on varying user characteristics (e.g., demographics, awareness, comfort), privacy settings, security measures, trust levels, and contextual harms and risks factors. By integrating the MPT, the systems could explore and predict various possible outcomes, each representing a different universe of potential decisions and consequences.

Contribution. With MPT, this paper contributes a novel lens for privacy modeling, grounded in the assumption that user privacy experiences are not singular but manifold. The theory offers:

- A probabilistic understanding of privacy under uncertainty.
- An explainable framework for personalized privacy-related decisions.
- A bridge between CI and empirical metrics (e.g., utility, trust, risk).

The ultimate goal is modeling the various potential states of privacy, risk, and trust, and how these states evolve over time to optimize decisions for the user. MPT also aims to balance data governance strategies, personalize privacy controls, and analyze AI system behavior under varying sociotechnical conditions.

2 The Multiverse Privacy Theory

To formalize the multiverse-inspired approach to privacy decision-making over time, we define a model in which each user action leads to a set of possible universes reflects a different scenario shaped by both user and system context and choice.

Definition. Let at each time step t, a set of possible universes $\mathcal{U}_t = \{U_t^1, U_t^2, \ldots, U_t^n\}$ arises from the user's selected privacy action $a_t \in \mathcal{A}_t$, where $\mathcal{A}_t$ denotes the available privacy actions at that time. Each universe reflects how varying contextual factors and the chosen privacy action influence outcomes. The probability of a universe U_t^i occurring, given privacy action a_t and context C_t, is modeled as:

$$P(U_t^i \mid a_t, C_t) \tag{1}$$

The utility of taking a privacy action a_t in context C_t is defined as:

$$\mathrm{UI}(a_t, C_t) = \alpha \cdot \rho_t + \beta \cdot S_t - \gamma \cdot R_t + \delta \cdot T_t + \zeta \cdot g(D_t) + \theta \cdot \mathrm{CI}(a_t, C_t) \tag{2}$$

where:

- ρ_t: User's privacy preference at time t
- S_t: Security level of the system/environment at time t
- R_t: Risk level faced by the user in the current context
- T_t: User's trust in the system at time t
- $g(D_t)$: Influence of demographic attributes such as age, sexual orientation, political affiliation, and familiarity with technology, etc.
- $\mathrm{CI}(a_t, C_t)$: CI score of the privacy action in the current context

Coefficients $\alpha, \beta, \gamma, \delta, \zeta, \theta$ are tunable weights representing the relative importance of each component in the utility function. To choose the optimal privacy action a_t^*, the system seeks to maximize the expected utility over all possible universe outcomes:

$$a_t^* = \arg\max_{a_t \in \mathcal{A}_t} \mathbb{E}[\mathrm{UI}(a_t, C_t)] \tag{3}$$

where the expected utility is defined as:

$$\mathbb{E}[\mathrm{UI}(a_t, C_t)] = \sum_{i=1}^{n} P(U_t^i \mid a_t, C_t) \cdot \mathrm{UI}(a_t, C_t, U_t^i) \tag{4}$$

Here, $\mathrm{UI}(a_t, C_t, U_t^i)$ denotes the utility of action a_t considering the outcome in universe U_t^i, possibly adjusted for universe-specific consequences. In addition, MPT introduces a recursive component to account for long-term consequences of privacy actions. The value function at time t is given by:

$$V_t = \mathbb{E}[\mathrm{UI}(a_t, C_t)] + \lambda \cdot V_{t+1} \tag{5}$$

where $\lambda \in [0,1]$ is a discount factor representing the weight given to future utility relative to the present. The recursive formulation enables the system to adapt and optimize privacy actions over time as user preferences and contextual factors evolve.

Example. To illustrate MPT, we implemented a Monte Carlo-style simulation across 5 distinct universes, each representing a plausible configuration of user preferences and contextual factors. Each universe U_i was simulated over 10 time steps. At each time step t, values were randomly generated to represent privacy preferences, security level, contextual risk, trust, and demographic sensitivity. Each privacy decision at a given time step resulted in multiple possible outcomes, reflecting alternative universes influenced by these factors. A utility value was computed for each outcome using a weighted combination of the input variables, with weights set to 1.0 for privacy preference (α), 0.8 for security level (β), -0.9 for contextual risk (γ), 0.6 for trust (δ), and 0.5 for demographic sensitivity (ζ). Additionally, to quantify how well a privacy decision aligns with favorable outcomes in the current context, we computed a CI score for each privacy action at time t as:

$$CI_i(t) = \frac{P_i(t) + S_i(t) + T_i(t) + D_i(t)}{1 + R_i(t)} \tag{6}$$

where $P_i(t)$, $S_i(t)$, $T_i(t)$, $D_i(t)$, and $R_i(t)$ represent normalized values for privacy preference, security level, trust, demographic sensitivity, and risk, respectively. The CI score increases with higher privacy preference, security, trust, and demographics, and decreases with increasing contextual risk.

At each time step, the system evaluates all possible privacy actions by calculating the expected utility over all universes and selects the action that maximizes this expected utility. Utility values were tracked over time and analyzed across three contextual risk bands: low, moderate, and high. Figure 1 illustrates the evolution of utility across the 10 time steps, showing how contextual risk influences privacy-related decisions.

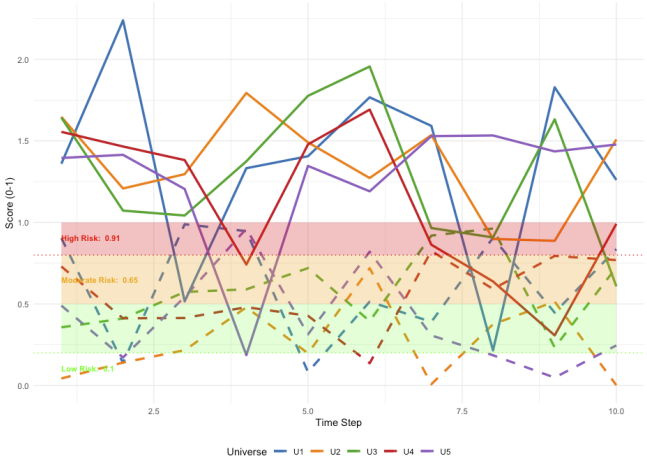

Figure 1: Simulation showing how user privacy utility evolves across multiple universes and three contextual risk bands.

Hypotheses Testing. We also tested the relationships between different factors and the utility score using Pearson's correlation. The results are summarized as follows:

H1: Privacy preferences significantly affect utility. A strong positive correlation was found between privacy preference and utility ($r = 0.6618$, $p < 0.001$), indicating that users who value privacy more tend to derive greater utility from privacy-related decisions. This result supports the hypothesis and is statistically significant.

H2: Higher contextual risk reduces privacy utility. There is a strong negative correlation between contextual risk and utility ($r = -0.6078$, $p < 0.001$). This indicates that as contextual risk increases, perceived utility decreases. The result is statistically significant and supports the model's assumptions.

H3: Trust increases privacy utility. A moderate positive correlation was observed between trust and utility ($r = 0.4407$, $p < 0.01$). This result is statistically significant and supports the hypothesis that trust positively influences privacy-related satisfaction.

H4: Security level does not significantly affect privacy utility. The correlation between security level and utility was weak and not statistically significant ($r = 0.1581$, $p = 0.2728$). This suggests that, in this model, security level alone may not directly impact perceived utility. One possible explanation is that users' perceived utility depends more strongly on direct privacy preferences, risk, and trust rather than on abstract or technical security metrics. It is also possible that users may lack sufficient understanding of security levels. However, this needs further exploration.

H5: CI strongly enhances privacy utility. A very strong positive correlation was found between CI and utility ($r = 0.8129$, $p < 0.001$), supporting the hypothesis that privacy decisions aligned with contextual expectations yield high utility.

By considering these factors, the system can simulate multiple potential scenarios across a wider range of privacy preferences, security concerns, and contextual influences. In other words, MPT can address evaluating these different universes by incorporating user demographics into each simulation, ultimately determining the optimal privacy decisions for each individual user. These decisions are made dynamically and adjust as the user's preferences evolve over time, leading to a personalized privacy experience that is adaptive and context aware. In addition, in this model, a system can continuously learn and adjust its decisions based on changing user demographics, ensuring that the best privacy practices are applied in ways that align with the user's preferences, context, and trust level. This process is inherently probabilistic, as the

Table 1: Summary of Hypothesis Testing Results

Hypothesis	MPT Attributes	r	p-value	95% CI
H1	Privacy Preference – Utility	**0.6618**	**1.668e-07**	[0.4700, 0.7939]
H2	Contextual Risk – Utility	**-0.6078**	**2.855e-06**	[-0.7579, -0.3965]
H3	Trust – Utility	**0.4407**	**0.0014**	[0.1871, 0.6380]
H4	Security Level – Utility	0.1581	0.2728	[-0.1258, 0.4181]
H5	CI – Utility	**0.8129**	**7.544e-13**	[0.6908, 0.8899]

system must choose between several possible universes of future interactions. By evaluating multiple potential universes, the system can mitigate risks of data exposure and maintain a balance between user privacy and security across various contexts.

3 Conclusion and Future Work

This paper introduces a novel MPT approach to privacy decision-making, considering the dynamic, evolving nature of privacy risks and user preferences. By simulating an infinite set of universes and using a time-dependent recursive model, AI systems can continuously optimize privacy settings in response to changing contexts.

Future work will explore MPT to test it to the real-world applicability of the model with practical deployment constraints and user expectations. We will explore (1) how to handle real-time threats, (2) refine the model with user demographics and contextual factors, and (3) incorporate regulatory compliance changes. To do this, we will conduct multiple survey design studies including demographic and scenario-based variations to reasonably and statistically evaluate infinite universes with real-world AI user populations. These scenarios will incorporate persona-specific risk models and adaptive utility functions.

References

[1] Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, and Nick Feamster. Discovering smart home internet of things privacy norms using contextual integrity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):1–23, 2018.

[2] Michaela Benk, Léane Wettstein, Nadine Schlicker, Florian von Wangenheim, and Nicolas Scharowski. Bridging the knowledge gap: Understanding user expectations for trustworthy llm standards. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(2626):27197–27205, April 2025.

[3] Abhilash Chakraborty, Anupam Biswas, and Ajoy Kumar Khan. Artificial intelligence for cybersecurity: Threats, attacks and mitigation. In *Artificial intelligence for societal issues*, pages 3–25. Springer, 2023.

[4] Jake Chanenson, Madison Pickering, and Noah Apthrope. Automating Governing Knowledge Commons and Contextual Integrity (GKC-CI) Privacy Policy Annotations with Large Language Models. *Proceedings on Privacy Enhancing Technologies*, 2025.

[5] Danielle Keats Citron and Ari Ezra Waldman. Rethinking youth privacy. *Virginia Public Law and Legal Theory Research Paper*, (2025-15), 2025.

[6] David Deutsch. Apart from Universes. *Many Worlds*, pages 542–552, 2010.

[7] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[8] Fabiano Damasceno Sousa Falcão and Edna Dias Canedo. Investigating Software Development Teams Members' Perceptions of Data Privacy in the Use of Large Language Models (LLMs). In *Proceedings of the XXIII Brazilian Symposium on Software Quality*, pages 373–382, 2024.

[9] Alisa Frik, Leysan Nurgalieva, Julia Bernd, Joyce Lee, Florian Schaub, and Serge Egelman. Privacy and Security Threat Models and Mitigation Strategies of Older Adults. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, page 21–40, 2019.

[10] Christine Geeng, Mike Harris, Elissa Redmiles, and Franziska Roesner. " like lesbians walking the perimeter": Experiences of {US}.{LGBTQ+} folks with online security, safety, and privacy advice. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 305–322, 2022.

[11] Ece Gumusel. A literature review of user privacy concerns in conversational chatbots: A social informatics approach: An Annual Review of Information Science and Technology (ARIST) paper. *Journal of the Association for Information Science and Technology*, 76(1):121–154, 2025.

[12] Maanak Gupta, CharanKumar Akiri, Kshitiz Aryal, Eli Parker, and Lopamudra Praharaj. From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access*, 11:80218–80245, 2023.

[13] Irit Hadar, Tomer Hasson, Oshrat Ayalon, Eran Toch, Michael Birnhack, Sofia Sherman, and Arod Balissa. Privacy by designers: software developers' privacy mindset. *Empirical Software Engineering*, 23:259–289, 2018.

[14] Yunhan Jack Jia, Qi Alfred Chen, Shiqi Wang, Amir Rahmati, Earlence Fernandes, Zhuoqing Morley Mao, Atul Prakash, and SJ Unviersity. Contexlot: Towards providing contextual integrity to appified iot platforms. In *ndss*, volume 2, pages 2–2. San Diego, 2017.

[15] Priya Kumar, Shalmali Milind Naik, Utkarsha Ramesh Devkar, Marshini Chetty, Tamara L. Clegg, and Jessica Vitak. "no telling passcodes out because they're private": Understanding children's mental models of privacy and security online. *Proceedings of ACM Human-Computer Interaction*, 1(CSCW):64:1–64:21, December 2017.

[16] Nora McDonald and Andrea Forte. Privacy and vulnerable populations. In *Modern Socio-technical Perspectives on Privacy*, pages 337–363. Springer International Publishing Cham, 2022.

[17] Michael B Mensky. Phenomenology of "dark matter"-from the everett's quantum cosmology. *arXiv preprint arXiv:1105.3696*, 2011.

[18] Helen Nissenbaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, Stanford, 2009.

[19] Ayodeji Oseni, Nour Moustafa, Helge Janicke, Peng Liu, Zahir Tari, and Athanasios Vasilakos. Security and privacy for artificial intelligence: Opportunities and challenges. *arXiv preprint arXiv:2102.04661*, 2021.

[20] Ghadeer Ghazi Shayea, Mohd Hazli Mohammed Zabil, Mustafa Abdulfattah Habeeb, Yahya Layth Khaleel, and A. S. Albahri. Strategies for protection against adversarial attacks in ai models: An in-depth review. *Journal of Intelligent Systems*, 34(1), January 2025.

[21] Yan Shvartzshnaider, Madelyn Rose Sanfilippo, and Noah Apthorpe. GKC-CI: A unifying framework for contextual norms and information governance. *Journal of the Association for Information Science and Technology*, 73(9):1297–1313, 2022.

[22] Aleksandra Slavković and Jeremy Seeman. Statistical data privacy: A song of privacy and utility. *Annual Review of Statistics and Its Application*, 10(1):189–218, 2023.

[23] Zhiping Zhang, Michelle Jia, Hao-Ping Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–26, 2024.

[24] Michael Zimmer. Addressing conceptual gaps in big data research ethics: An application of contextual integrity. *Social Media+ Society*, 4(2):2056305118768300, 2018.