Received: 5 November 2022 | Revised: 23 December 2022 | Accepted: 26 December 2022 | Published online: 27 December 2022

RESEARCH ARTICLE

Artificial Intelligence and Applications
2023, Vol. 1(2) 82–89
DOI: 10.47852/bonviewAIA2202514

BON VIEW PUBLISHING

A Multi-Party Agent for Privacy Preference Elicitation

Rim Ben Salem[1,*], Esma Aimeur[1] and Hicham Hage[2]

[1]*Department of Computer Science and Operations Research, University of Montreal, Canada*

[2]*Science Department, Notre Dame University-Louaize, Lebanon*

Abstract: In today's world, the decisions that individuals make online often include their surroundings and social circles. For example, Alice posts on TikTok to celebrate her friend Bob's birthday and reminisce about their best memories together. She, then, proceeds to create a campaign to fund her local place of worship and tags members of her community who share her religious belief. Alice might equally like to take initiative at work as she plans her team-building trip and excitedly shares the programme on Facebook. While doing all of this, she is involving family members, close friends, co-workers, acquaintances, and others from her social circle, all of whom might have different opinions about their privacy. While she sees no issue with her actions, her friend Bob, for one, might not agree, hence, the issue of multi-party privacy. Many researchers have focused on conflict resolution, which occurs when the sharer's privacy preferences do not align with the other parties involved. However, one key point in this approach is eliciting the preferences of these individuals. Oftentimes, there is an underlying assumption that the system has sufficient historical data to represent the perspective of the multi-party members. The problem is that this is not always the case in real life and the cold start problem might be unavoidable. The system that is meant to nudge the sharer to reduce the multi-party disclosure might not even be capable of representing the preferences of everyone involved at the beginning. Hence, this paper addresses this issue through the use of the Classification and Regression Tree (CART) combined with the Rasch model. Study participants ($N = 800$) responded to realistic scenarios showcasing multi-party disclosure, which is used to construct and test the multi-party agent. The results suggest that the system performs well in overcoming the cold start problem as reported by the accuracy, precision, and recall.

Keywords: multi-party, privacy-preserving, eliciting, classification, Rasch model, cold start

1. Introduction

Social Networking Sites (SNS) have long since surpassed the constraints of the virtual universe and stepped into the real world. Decisions that users make online have implications for their day-to-day lives. There is no shortage of instances where Ivy League universities such as Harvard rescinded offers to students because of past controversial Tweets (Levin, 2020). Another example is that of a newly promoted employee who was discussing her salary increase on TikTok and got fired because her sharing behaviour caused her employer to distrust her (Diaz, 2022). The consequences of these decisions impact the individual's personal, professional, and social life alike. However, making the correct choice and behaving in the most beneficial way is becoming increasingly difficult. This is due to multiple reasons, one of which is the ever-changing landscape of IT in general and social media specifically. Let us take TikTok as a case in point, by the time it became popular, users started complaining after seeing their clips used to promote the app on other platforms without their permission (Hutchino, 2019). It was revealed that this is indeed stated in the user guidelines that TikTok is allowed to repost the submitted clips and profit from the likeability of the content creators in any way that is deemed fit. The surprise amongst users stems from a lack of knowledge concerning their rights and who owns the uploaded videos. Minaei et al. (2021) explain that sometimes deleting a post in retrospect makes users more vulnerable because malicious actors specifically signal this action as an intent to hide something that is damaging to the owner. Hence, providing preventive measures before the act is of the essence. One popular method to do this is through the use of nudges. A nudge is a mechanism of behavioral science that promotes an action in the best interest of the receiver. In the field of cybersecurity and privacy, it could be a pop-up with caution signs warning the user against proceeding with clicking on a suspicious link. There is a growing interest in adopting this approach to assist users, specifically, when they are about to disclose private information, which is self-disclosure.

However, there is a gap in the field of personalized privacy nudges that can handle *multi-party* (Such and Criado, 2018) conflicts. This term refers to "*co-owned*" data that involves people other than the sharer. A basic setting to illustrate possible multi-party privacy conflict would be if user Alice shares a photo with the public audience and tags her friend Bob acting funny after drinking. Bob is a private person and only shares personal content with close friends and family, hence, the issue with his friend's

*Corresponding author: Rim Ben Salem, Department of Computer Science and Operations Research, University of Montreal, Canada. Email: rim.ben.salem@umontreal.ca

post. Alice's post can have consequences on his professional life, and she might not have been aware of it. If Alice had been nudged before proceeding with the disclosure, Bob's preferences would have been respected and the issue would have been prevented. Although some scholars have proposed methods to mediate between the sharer and the multi-party members (such as voting on the best solution), they mostly assume that all the preferences are already known to the system. But, in real life, that is a strong assumption to make as this would require direct preference elicitation with each one of them. Generally, this is done through a questionnaire that might include disclosure scenarios to gauge their response and assign values to their user model. Thus, this raises the cold start problem in which the system does not have an initial value for the preferences of these multi-party members.

This paper aims to tackle these issues by predicting the multi-party members' preferences, drive for disclosure, and motivations when a direct elicitation is unfeasible (they have not interacted with the system before). We propose a novel multi-agent nudge-based system aimed at social media users in order to mitigate multi-party disclosure. This work contributes the following:

- Proposing a multi-agent system to represent all the parties involved in the disclosure.
- Establishing a user model encompassing the notion we define as the disclosure appetite as well as the context-specific disclosure goals.
- Focalizing on the multi-party agent, which is a component of the multi-agent system. We focus on the preference elicitation process in order to construct the multi-party user model.
- Evaluating the proposed system with the help of 800 participants who were remotely recruited over a period of 4 days using Amazon's Mechanical Turk and covering 3 geographical areas: "North America," "Europe," and "Asia."

The paper is organized as follows: Section 2 discusses existing work that relates to our research. Section 3 details our proposed system and the different submodules that constitute it. Section 4 reports on the evaluation of the system. Section 5 concludes this work and provides pointers to future works.

2. Literature Review

The National Institute of Standards and Technology (NIST) of the United States (McCallister, 2010) and the European Union's (EU) General Data Protection Regulation (GDPR) (General Data Protection Regulation, 2022) defined any physically unique, psychologically expressive, cultural, social, biometric, genetic and health data as Sensitive Personal Data (SPD) or Personally Identifiable Information (PII). There is a growing body of research focusing on protecting PII and identifying the risks associated with mishandling such information intentionally or unintentionally as stalking, identity theft, price discrimination, or blackmailing (Gross and Acquisti, 2005). Boyd and Ellison (2007) point out privacy risks such as damaged reputations, unwanted contacts, harassment, and use of personal data by third parties.

These risks extend beyond the individual who made the decision to share the privacy-jeopardizing content. Alice might have no qualms about sharing her own location when travelling, but if she does so on her team-building trip, her co-workers can be subjected to the consequences. Hence, the multi-party privacy dilemma arises. Solving or at least mitigating the problem requires a representation of not only the sharer but also the multi-party member. This hinges upon the user modelling process, which many scholars have approached in various ways. Omarzu proposed a disclosure decision model (Omarzu, 2000) to predict the disclosure decision in real-life scenarios. An example of this is as follows that romantic settings are disclosure catalysts when compared with mundane, less intimate scenarios such as being in a professional environment. In a similar context of user decision-making, people evaluate the risks and the perceived gratification (Dienlin and Metzger, 2016) and, depending on the situation, one can outweigh the other resulting in the decision to disclose or not the data. Going beyond representation and understanding to the actual conflict resolution, Krol and Preibusch (2015) coined *effortless privacy negotiations*. Their work aims to converge the heterogenous privacy preferences of the individual users involved to reach an agreement. Various techniques have been utilized for the same purpose and they range from question-based profiling (Hutton and Henderson, 2015; Norval and Henderson, 2019), ontology-based (Kökciyan et al., 2017), to game theoretic approaches (Yassine and Shirmohammadi, 2009).

All of these models and approaches depend on the user modelling and specifically the preference elicitation process. The easiest and most used way to achieve this has been through direct interactions. The most basic form of which is simply asking "how sensitive do you think your location is on a scale of 1-5 going from least to most sensitive?" (Ben Salem et al., 2021), for example. Although this can be criticized for potential biases resulting from the user self-evaluating their preferences (Lahtinen et al., 2020), a user model can be established and later updated. However, this is not the case for multi-party party privacy. The sharer can be a user of the system (who will eventually receive the nudge), but there is no guarantee that the other party members are the same. If they are not users, then how can the preference elicitation and user modelling be achieved? Moreover, the answer to that question needs to meet specific criteria as has been demonstrated in (Pu and Chen, 2008): a good elicitation strategy should increase prediction accuracy with minimum user interaction.

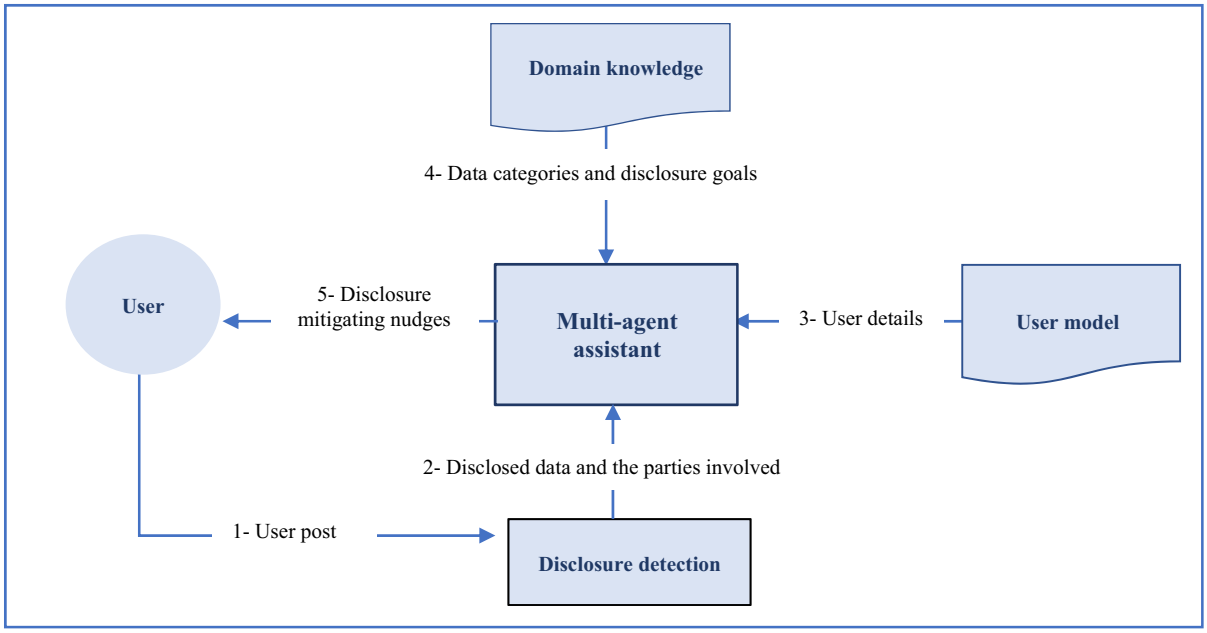
Our proposition aims to contribute to the privacy preference elicitation of multi-party members. The proposed nudge-based system as a whole uses multiple agents to represent the disclosure gain and privacy loss as perceived by each party, be it the user or others involved (sharer and others). The novelty resides in the fact that it addresses the cold start problem that often occurs when multi-party members cannot be directly questioned to establish their preferences. Moreover, the proposed solution, which consists of the multi-party agent, aims to limit the involvement of the sharer so as to not burden them.

The next section focuses on the nudge-based system in which resides the multi-party agent.

3. Nudge-Based System for Multi-Party Representation

The goal of the whole system is to mitigate disclosure on the multi-party level and intervene with a nudge that the sharer receives. It first starts by building a user model that reflects the goals and drives for disclosure that we introduce as our novel concept called: the *disclosure appetite*. This section will go through the different components of the system as seen in Figure 1 and arrive at the multi-party agent. The next subsection is dedicated to the domain knowledge.

Figure 1
Multi-agent system for disclosure mitigation on social media



3.1. Domain knowledge

The domain knowledge that this nudge-based system is based on is derived from the existing research centred around 2 main axes: first, the categories of personal data that the system deals with and second, the disclosure goals as perceived by the individual.

3.1.1. Categories of personal data

The term personal data refers to any information relating to an identifiable person. The *Social Penetration Theory* (SPT) (Taylor, 1968) uses an onion metaphor to detail categories of personal data that constitute what is known as "personality."

The onion is used metaphorically to describe sequentially removable layers that conceal an important something, which is privacy in the current context. This work draws inspiration from it and proposes a new updated perspective. Throughout this paper, when mentioning personal data, **ten categories** in total are considered, of which:

Three are drawn from the SPT: Biographical data (e.g. Name and age), Goals and fears (e.g. Ambitions and dreams), and religious and political convictions (e.g. political party and frequented place of worship). **Seven** are drawn from Personally Identifiable Information (PII): Medical records (e.g. Allergies and long-term afflictions), banking information (e.g. Transaction details and account login), diplomas/certificates (e.g. Scanned versions of official diplomas), official documents (e.g. ID and passport), photos (e.g. Selfies), most frequented locations (e.g. Favourite restaurant and park), travel plans (e.g. Accommodation plans and sightseeing).

3.1.2. Disclosure goals

Knowing that a piece of information can identify them and still disclosing it suggests a trade-off in the user's mind known as the privacy calculus (Trepte et al., 2017). Deducing what the user's aim is and how much it is worth to them is an indispensable process to the nudge generation that comes afterwards. In this work, four goals inspired by Aïmeur et al. (2020) as the authors identify the motivations and goals behind disclosing personal data:

Financial gain: The motivation is monetary gain.
It can take the form of cash, digital or virtual currency.

Personal gain*:* This encompasses all nonmonetary services such as exclusive access to premium services.

Moral gain/altruism*:* The user who aims to achieve an altruistic goal motivated by a sense of morality and virtue.

Social compliance*:* This goal encompasses Cialdini's *principles of persuasion* (Cialdini, 2007). At its core, his work details how peoples' decisions are highly influenced by their surroundings and relationships within their social circles.

The domain knowledge is used as the basis for the user model, which is the foundation of the multi-agent assistant.

3.2. User model

There are 2 user models needed to push nudges to the sharer in the context of multi-party disclosure: the sharers and the multi-party members'. We consider 2 parameters: disclosure appetite and the perceived data sensitivity.

3.2.1. Disclosure appetite for multi-party members

In the context of enterprise risk management, the term "*risk appetite*" has a number of definitions, most with a link to risk acceptability, but also values and goals. It is one of the decisive parameters in decision-making and is often used interchangeably with "risk acceptability" and "risk tolerability." This article borrows that term and adapts it into "*disclosure appetite*" to fit the context of the disclosure, both self-disclosure and multi-party disclosure. Essentially, the disclosure appetite can be seen as the social media user's acceptable level of privacy compromise in order to seek a specific goal. An example of this is: The user Bob

does not like to express vulnerability online, but he might not mind it when his friend Alice and others are sharing sad stories involving him. This is an expression of the "social compliance" goal (discussed in the domain knowledge subsection). As a result, his disclosure appetite can be high in this context, but not in others like exchanging his privacy for monetary gain. However, Bob's preferences are not solely contingent on the goal, but also depend on who is disclosing his information. If we consider the same scenario, he is okay with Alice, his friend, committing the disclosure, but not necessarily his co-worker.

With this in mind, 3 components of context are considered:

- The connection of the multi-party member to the sharer: Is the other person tagging Bob a close friend/family member, a co-worker/classmate or a stranger from the general public?
- The audience: Who has access to the shared information? Is the photo shared with friends or is it available to anyone? The audience is also divided into 3 social circles: *close friends and family, co-workers/classmates,* and *the public.*
- The goal of the disclosure: As detailed in the domain knowledge, there are numerous goals and while one social media user might be moved by financial gain the most, another can be motivated by social compliance. Next is the perceived data sensitivity.

3.2.2. Perceived data sensitivity

Data valuation and privacy, in general, are subjective and the user's background shapes their preferences, estimations, and opinions. For each user, we propose to represent the way they perceive their privacy in a layered structure inspired by the *Social Penetration Theory* (SPT) as seen in Figure 2. SPT proposes a six-layer model to order data based on lowest to highest sensitivity: *Biographical data, preference in clothes, food and music, goals and aspirations, religious convictions, deeply held fears and fantasies and the most sensitive being: concept of self.*

Table 1
An excerpt of Bob's user model (multi-party member) for one data category

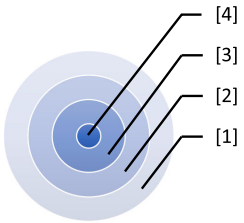
	Relationship with the sharer		
	Friends/family	Co-workers/ classmates	General public
Disclosure appetite	0.7	0.4	0.1
Perceived sensitivity	0.3	0.8	0.2

Table 1 shows a part of a user model associated with one specific piece of data: the values of disclosure appetite and perceived sensitivity in each context. This example can be interpreted as follows: Bob's disclosure appetite when the piece of data "location" is being shared by a family member or his close friend Alice is 0.7. It drops to 0.4 if the sharer is his co-worker. The perceived data sensitivity, on the other hand, increases from 0.3 if the sharer is a friend/family member to 0.8 if it is a co-worker. If "location" is replaced by "religious and political belief," these values are different. Hence, why in total, each user model has 30 disclosure appetite values and 30 perceived sensitivity values. This corresponds to 10 categories of personal data as defined in the domain knowledge multiplied by 3 social circles that the sharer can belong to. This concludes the user model, which is the basis of the multi-party agent.

3.3 Multi-agent assistant

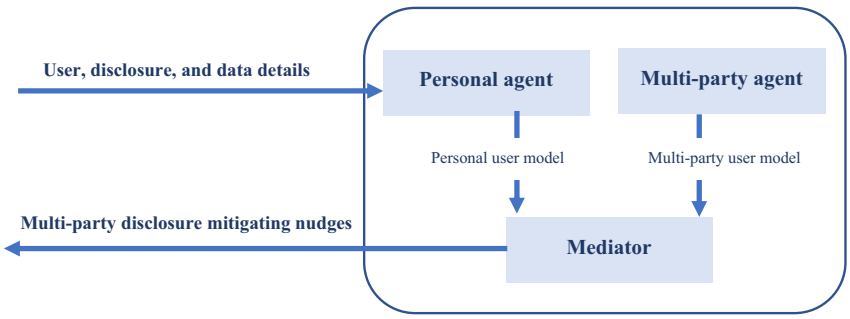
The assistant is the core of the system, which is designed to mitigate multi-party disclosure. It is made of 3 agents as seen in Figure 3. It is the nudge-generating main component that takes in

Figure 2
Social penetration-inspired classification of personal data



1. Biographical data
2. Preferences, Selfie, charity contribution
3. Travel plans, religious and political opinions, license plate, photos of diploma
4. Banking information, official documents (ID, passport), Medical records, home address

Figure 3
Multi-agent assistant



the information being disclosed, and the individuals' preferences and then, comes up with the best context-specific nudge for the situation.

3.3.1. Personal agent

The personal agent represents the user's preference, namely their disclosure appetite and data sensitivity as described in the user model section. If there is a case of multi-party disclosure involving the sharer and another party, the personal agent considers the following:

- The user's drive and expected gain by disclosing their own data: disclosure appetite
- The user's drive and expected gain by disclosing the other party's data: multi-party disclosure appetite
- The user's perceived sensitivity of their own data: data sensitivity
- The user's perceived sensitivity of the other party's data: multi-party data sensitivity

All of these values are context-specific; the disclosure appetite associated with sharing a selfie for financial gain is different from the same metric when the goal is social compliance or aligning with an altruistic belief, etc. The next section elaborates on the multi-party agent.

3.3.2. Multi-party agent

Not every person involved in the disclosure undergoes the direct preference elicitation process in order to construct the user model. This is needed when the user Bob is tagged in a post by someone else, namely his friend Alice. While the personal agent explains Alice's motivation for sharing from her personal perspective, this section considers the perspective of Bob who needs a representative due to his involvement. It is worth noting that if the multi-party privacy disclosure involves more than one individual, each of them will be represented by an agent to ensure their best interests as well. When the sharer links/tags another person such as Bob, the user database is searched for a matching profile. There are 2 possible scenarios when such a thing happens:

- *Scenario 1: Bob is a user who is already recognized by the system and has a user model that includes his disclosure appetite and goals.*
- *Scenario 2: Bob is either completely unknown to the system or his user model is missing some values.*

Scenario 1 does not require much detailing, since the system already knows the user's privacy needs and preferences. Indeed, in this case, an agent can be simply assigned to Bob in accordance with his already available data. Scenario 2, however, requires the system to predict the preferences of Bob. In order to do so, the system relies on 2 inputs:

- The sharer Alice's input. She is asked a few questions to estimate Bob's likely preferences.
- Existing users that the system recognizes who resemble Bob's user model. Other individuals are considered similar to Bob based on Alice's input and what can be detected from the post.

There are a few considerations to take into account. First, the fact that the sharer Alice should not be burdened by asking her many questions about Bob otherwise she might abandon the process. So, amongst all the demographic data (gender, age, education, occupation, ethnicity), and questions such as "Is Bob often public about his political opinion," which to choose and ask Alice about?

Second, this depends on the studied population and the use of a predictive modelling method. We do not have Bob's information, but we do have the user model of other users like Alice.

The proposed approach uses this dataset to indirectly elicit Bob's preferences. This relies on a combination of CART and the *Rasch model.*

The use of CART: Figure 4 illustrates a small part of the actual regression tree, built based on the data we gathered (further information on the data sample and the collection process is detailed in the evaluation). CART is one of the most widely used algorithms for training axis-aligned decision trees.

It applies a greedy recursive partitioning, which optimizes a pure measure (Gini index) at each node. When splitting a given node, it enumerates all the features and thresholds to find the split that maximally reduces the Gini index. It continues to grow a tree up to a maximum depth and then prunes nodes, one by one until a cost-complexity criterion is met. The pruning process eliminates the less significant nodes, which in return reduces the number of questions that Alice needs to answer in order to predict Bob's preferences and overcome the cold start problem. Following the tree presented in Figure 4, the first question to ask the sharer (Alice) is the age, and then depending on the answer, we navigate the branches until the leaf where the value resides in. So, for the age 25-34 and with an occupation of a researcher, Bob is estimated to have a disclosure appetite $= \boldsymbol{\beta 1}$.

However, one question remains: How to use the disclosure appetite and the perceived data sensitivity of the multi-party members to predict their preferred outcome (to disclose or not)? To achieve this 2-parameter-based balance, we choose to work with Item Response Theory (IRT) as this paradigm, by design, is user-centric and fits our design.

The use of the Rasch model: In psychometrics, item response theory is a paradigm for the design, analysis, and scoring of tests and questionnaires. A special case of IRT is *the Rasch model*, which takes as input "*user ability*" and "*item difficulty*" and the output is the probability of them answering the question correctly. Our proposed system aligns with this duality of "personal parameter" (disclosure appetite) and "item parameter" (data value) and as such, we propose an adapted version of the Rasch model. In short, the original Rasch model calculates the probability of a person answering an item correctly, based on that person's ability level and the difficulty of the item. In our context, we adapt the model by substituting the ability of the person with the disclosure appetite, and the difficulty of the item with the sensitivity of the data. Such an approach allows the system to determine the probability of the multi-party member agreeing with the disclosure of the sharer.

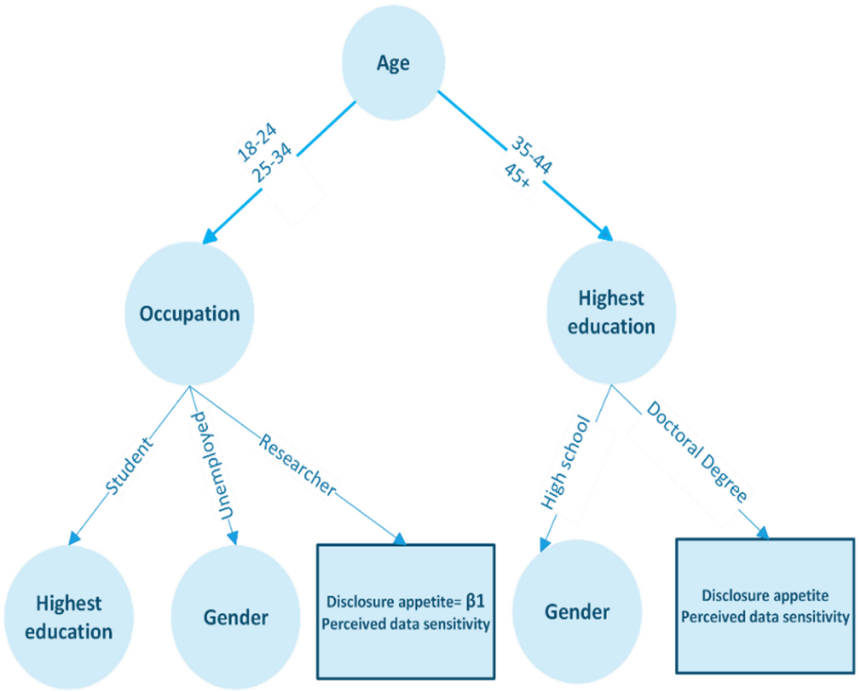
In the context of the proposed nudge system, $X_{ni} = x \in \{0, 1\}$ is the dichotomous random variable where $x = 1$ denotes an accepted disclosure by multi-party member n concerning item i. P is the probability of the opposite outcome $X_{ni} = 1$:

$$P\{X_{ni} = 1\} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}} \quad (1)$$

where: βn is user-specific disclosure appetite and δi is sensitivity/ value of a piece of data.

To explain this, we consider two actors Alice and her co-worker and the disclosure goal "moral gain," for example. Alice is the sharer, and the co-worker is the multi-party member. To make the example simpler, we fix the sensitivity value for both of them at

Figure 4
A snippet of a CART to estimate the multi-party user model parameters



0.21, and we attribute different disclosure appetite measures: Alice's = 0.87 and co-worker's = 0.11.
For Alice:

$$P\{X_{ni}=1\} = \frac{e^{\beta_n-\delta_i}}{1+e^{\beta_n-\delta_i}} = \frac{e^{0.87-0.21}}{1+e^{0.87-0.217}} = 0.66$$

For Bob:

$$P\{X_{ni}=1\} = \frac{e^{\beta_n-\delta_i}}{1+e^{\beta_n-\delta_i}} = \frac{e^{0.11-0.21}}{1+e^{0.11-0.21}} = 0.47$$

The decision to share aligns with Alice's preferences but not her co-worker's ($P<0.5$). This concludes the multi-party agent and the remaining agent is the mediator.

3.3.3. Mediator

The purpose of the mediator is to consider all parties involved in the disclosure and make personalized nudges to the user. At this point, considering the personal, and multi-party agents, the system can proceed to the conflict resolution process. Aggregation-based approaches (Carminati and Ferrari, 2011), for example, are designed so that each agent can cast a vote and the final outcome is determined by the majority. A special case of this would be veto voting (Thomas et al., 2010) in which each user is allowed to oppose sharing and unless a unanimous agreement is reached, the nudge pushed to the sharer would be against sharing the content. Other mediation approaches include auction-based systems (Squicciarini et al., 2009) where users gain fictitious money that they can invest in auctions. It is basically a bidding for the most desired sharing decision for co-owned items.

4. Evaluation

For the evaluation, *800 people* were remotely recruited over a period of 4 days using Amazon's Mechanical Turk. It covered 3 geographical areas: "North America," "Europe" and "Asia."

The survey has *60 questions,* and it is estimated to take participants *15-20 minutes* to complete it. It starts with a consent form providing the purpose of the research, affiliation of the researchers, and information on the anonymity of the responses and the right of withdrawal. Moreover, information on the ethical board that has approved this research is provided to the users with references to our specific project. The survey questions are separated into 3 parts as follows:

General demographic questions: age, gender, highest education, current occupation, and origin. Such information is necessary to help construct the CART model whose nodes correspond to demographical details.

Disclosure appetite questions: Participants are provided a 1-10 scale to record their responses, and they are asked questions pertaining to how agreeable they are with some disclosure scenarios. An example would be "Would you be okay if your friend shares your vacation plans publicly on social media?"

Perceived data sensitivity questions: Participants are asked "how sensitive do you think this piece of data is." They are provided with a 1–10 scale to record their response.

The evaluation uses a 5X2-fold cross-validation. We rely on the goodness of fit to report on the performance of CART as seen in Table 2. It is a useful way to identify the discrepancy between observed values and the values expected under the model in question. The results are promising, but this is solely an indicator of how well the model attributes the disclosure appetite and the data sensitivity initially.

Table 2
Goodness of fit of CART

Metric	Accuracy	Recall	Precision	F1-score	Specificity	Sensitivity
Value	0.82	0.85	0.83	0.83	0.78	0.85

Table 3
The performance of the multi-party agent

The sharer's social circle	Correctly classified instances	Mean Absolute Error (MAE)
Close friends and family	94%	0.08
Co-workers and classmates	91%	0.13
The public	65%	0.81

Another important aspect is how good the preference elicitation is using the Rasch model. To do this, we set aside 200 users out of the 800 we recruited. Their answers were not given to the system as if they were new users without any established user model. Then, we used CART to predict their user model and measured the probability of them being agreeable to various disclosure scenarios using the Rasch model. Finally, this predicted multi-party agent is put to the test by comparing it to the actual model established directly through the responses of users.

If the model predicts that the user would agree to the disclosure, but in reality, they are not (according to their response), this is a false positive. A false negative would be predicting a rejection when the user does not mind the disclosure. Table 3 reports on the performance of the agent depending on the social circle of the sharer. The correctly classified instances (true positives and true negatives) constitute 94% when the sharer belongs to "close family and friends." The lowest is 65% for the general public. We think that a potential reason for this is the fact that people are not very concerned about the privacy of the general public. They might be considerate to close friends and family members, but when it comes to complete strangers, they are not as interested in their privacy preservation.

Finally, we would like to point out a few limitations that we aim to tackle in the future. The *first* of which is that the 800 participants were recruited from North America, Europe, and Asia. A follow-up study needs to include more diversity based on the geographical location. *Second*, the evaluation, although promising, is based on realistic scenarios instigated by the system to which the participants respond. A real-life setting can prove to be more challenging. *Third*, our model hinges on the cooperation of the sharer by answering the relevant questions to elicit the preferences of the multi-party member. One way to reduce this task would be to infer as much information as possible from the content itself.

Implicitly eliciting the preferences of the multi-party members can further reduce the questions asked to the sharer and improve the user modelling process.

5. Conclusion

Providing a decision-making assistant for social media users to help them navigate the platforms is a priority in today's world. Individuals disclose too much information on a daily basis because they seek various goals. Furthermore, the issue goes beyond self-disclosure to include their social circles and people whose disclosure appetite might be very different from the sharer. This is the conflict that can arise involving multi-party privacy. The existing solutions to mitigate the issue are heavily reliant on the individual's disclosure preferences. When multi-party members are represented in multi-agent systems, the focus is often on the resolution rather than figuring out how to elicit their preferences.

Indeed, these individuals might be completely unknown to the system and a direct preference elicitation could be unfeasible. This paper addresses this cold start issue and proposes a multi-party agent to establish a preliminary user model for everyone involved in the disclosure. This is achieved through the combination of the Rasch model and CART. Predicting the preferences using the latter and using them as inputs for the Rasch model yields promising results (goodness of fit of CART and the overall performance of the multi-party agent) when evaluated on real users. The proof of concept of our multi-party agent is encouraging as the first step towards a real-life implementation in more social networks in which the sharers face situations requiring the intervention of the nudges to mitigate the disclosure involving others. Moreover, in the future we would like to focus on generalizing the model on a larger scale. Although we designed this process to be succinct and to reduce the engagement on the sharer's part, the task can become complex if the co-owned content involves many people. It would be interesting to look into the preference elicitation of multiple multi-party members.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

References

Aïmeur, E., Díaz Ferreyra, N., & Hage, H. (2020). Manipulation and malicious personalization: Exploring the self-disclosure biases exploited by deceptive attackers on social media. *Frontiers in Artificial Intelligence*, *2*, 26. https://doi.org/10.3389/frai.2019.00026

Ben Salem, R., Aïmeur, E., & Hage, H. (2021). The privacy versus disclosure appetite dilemma: Mitigation by recommendation. *In Workshop on online misinformation- and harm-aware recommender systems, ACM recommender systems conference.*

Boyd, D., & Ellison, N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Carminati, B., & Ferrari, E. (2011). Collaborative access control in on-line social networks. In *7th international conference on collaborative computing: Networking, applications and worksharing,* 231–240.

Cialdini, R. B. (2007). Influence: The Psychology of Persuasion (*Revision Edition*).

Dienlin, T., & Metzger, M. J. (2016). An extended privacy calculus model for snss: analyzing self-disclosure and self-withdrawal in a representative U.S. sample. *Journal of Computer-Mediated Communication*, *21*(5), 368–383. https://doi.org/10.1111/jcc4.12163

Diaz, A. (2022). I got fired for sharing my salary on TikTok — and cried for days straight. Retrieved from: https://nypost.com/2022/07/21/i-got-fired-for-sharing-my-salary-on-tiktok-and-cried-for-days-straight/

General Data Protection Regulation (2022). Retrieved from http://gdpr-info.eu

Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, 71–80.

Hutton, L., & Henderson, T. (2015). "I didn't sign up for this!": Informed consent in social network research. In *Proceedings of the International AAAI Conference on Web and Social Media. 9*(1), 178–187.

Hutchinso, A. (2019). TikTok's Turning User-Submitted Content into Ads, Without User Knowledge. Retrieved from: https://www.socialmediatoday.com/news/tiktoks-turning-user-submitted-content-into-ads-without-user-knowledge/564518/

Kökciyan, N., Yaglikci, N., & Yolum, P. (2017). An argumentation approach for resolving privacy disputes in online social networks. *ACM Transactions on Internet Technology*, *17*(3), 1–22. https://doi.org/10.1145/3003434

Krol, K., & Preibusch, S. (2015). *Effortless privacy negotiations. IEEE Security & Privacy*. *13*(3), 88–91.

Lahtinen, T. J., Hämäläinen, R. P., & Jenytin, C. (2020). On preference elicitation processes which mitigate the accumulation of biases in multi-criteria decision analysis. *European Journal of Operational Research*, *282*(1), 201–210. https://doi.org/10.1016/j.ejor.2019.09.004

Levin, D. (2020). Colleges rescinding admissions offers as racist social media posts emerge. *The New York Times*. Retrieved from: https://www.socialmediatoday.com/news/tiktoks-turning-user-submitted-content-into-ads-without-user-knowledge/564518/

McCallister, E. (2010). *Guide to protecting the confidentiality of personally identifiable information*. Diane Publishing.

Minaei, M., Mouli, S. C., Mondal, M., Ribeiro, B., & Kate, A. (2021). Deceptive deletions for protecting withdrawn posts on social media platforms. In *NDSS*.

Norval, C., & Henderson, T. (2019). Automating dynamic consent decisions for the processing of social media data in health research. *Journal of Empirical Research on Human Research Ethics*, *15*(3), 187–201. https://doi.org/10.1177/1556264619883715

Omarzu, J. (2000). A disclosure decision model: Determining how and when individuals will self-disclose. *Personality and Social Psychology Review*, *4*(2), 174–185. https://doi.org/10.1207/S15327957PSPR0402_05

Pu, P., & Chen, L. (2008). User-involved preference elicitation for product search and recommender systems. *AI Magazine*, *29*(4), 93. https://doi.org/10.1609/aimag.v29i4.2200

Squicciarini, A. C., Shehab, M., & Paci, F. (2009). Collective privacy management in social networks. In *Proceedings of the 18th international conference on World Wide Web*, 521–530.

Such, J. M., & Criado, N. (2018). Multiparty privacy in social media. *Communications of the ACM*, *61*(8), 74–81. https://doi.org/10.1145/3208039

Taylor, D. A. (1968). The development of interpersonal relationships: Social penetration processes. *Journal of Social Psychology*, *75*(1), 79–90. https://doi.org/10.1080/00224545.1968.9712476

Thomas, K., Grier, C., & Nicol, D. M. (2010). Unfriendly: Multiparty privacy risks in social networks. In *Privacy enhancing technologies*, 236–252. Springer Berlin Heidelberg.

Trepte, S., Reinecke, L., Ellison, N. B., Quiring, O., Yao, M. Z., & Ziegele, M. (2017). A cross-cultural perspective on the privacy calculus. *Social Media + Society*, *3*(1), 2056305116688035. https://doi.org/10.1177/2056305116688035

Yassine, A., & Shirmohammadi, S. (2009). An intelligent agent-based framework for privacy payoff negotiation in virtual environments. In *2009 IEEE Workshop on Computational Intelligence in Virtual Environments*, 20–25. IEEE

How to Cite: Ben Salem, R., Aimeur, E., & Hage, H. (2023). A Multi-Party Agent for Privacy Preference Elicitation. *Artificial Intelligence and Applications* 1(2), 82–89, https://doi.org/10.47852/bonviewAIA2202514