# Towards Enhanced Privacy-preserving Nudges

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

[1] *Department of Computer Science & Operations Research, University of Montreal, Montreal, QC, Canada*
[2] *Science Department, Notre Dame University-Louaize, Zouk Mosbeh, Lebanon*

**Abstract**

Information and communications technology (ICT) is proliferating exponentially and has surely become an intrinsic part of our daily lives. However, its fast-paced growth has brought upon multiple challenges amongst which are cybersecurity and privacy. While the technical aspects such as protocols and risk management measures are important, the human side is often overlooked despite being considered the weakest link.

With the progress in Artificial Intelligence and its subfields, many contributions have been made to the topic of user protection online. Namely, smart adaptive agents are today capable of drawing patterns, predicting behaviours and assisting the user in the decision-making process. This paper examines both the objective non-user-specific and the subjective personalized approaches. It provides arguments supporting each and sheds light on their drawbacks. Then, it proposes a hybrid solution for privacy-preserving AI-based agents, which has the potential to mitigate the discussed shortcomings.

**Keywords**

Privacy-preserving, risk management, adaptive agents, objective agents.

## 1. Introduction

Privacy is a fundamental issue for those involved in human-computer interaction. With the omnipresence of social media, privacy and cybersecurity risks, researchers began to identify potential privacy and safety risks [1]. Hence, there is a growing interest in designing ways to assist users to adopt safe behaviour online. Specifically, privacy-preserving nudges have garnered increasing attention in recent years [2] [3]. Nudges, in general, have been heralded as offering intervention and a "push" to promote desirable positive behaviours such as saving for retirement and charitable giving [4]. In the context of digital welfare, the applications of this mechanism include visualizing the strength of a password to encourage users to strengthen it [5] [6]. Mitigating self-disclosure on social networks is a popular application of cues [7] where soft behavioural reinforcements are applied to guide users. Kroll *et al.* [8] reported that reminders to change privacy settings trigger privacy concerns, which can result in positive behavioural changes, at least in the short term.

With the progress of Artificial Intelligence (AI) and its wide range of domains, multiple forms of smart intuitive behavioural reinforcements have been developed. Prompts, cues, and notifications are all forms of nudges, which can rely on general knowledge or user-specific parameters. The existing solutions have been garnering attention and are generally divided into two categories: preference-based personalized agents and objective neutral agents. The personalized agents rely on a user's disclosure preferences and goals to deliver nudges. On the other hand, the second type of privacy assistants does not model the user's behaviour nor does it try to customize the advice depending on the individual. We have worked on both types of assistants separately in our earlier

*AIofAI '22: 2nd Workshop on Adverse Impacts and Collateral Effects of Artificial Intelligence Technologies, Vienna, Austria*
EMAIL: rim.ben.salem@umontreal.ca (A. 1); aimeur@iro.umontreal.ca (A. 2); hhage@ndu.edu.lb (A. 3)
ORCID: 0000-0001-6938-0097 (A. 1); 0000-0001-7414-5454 (A. 2); 0000-0002-2267-9371 (A. 3)

research, which gives us insight into the subject. This paper examines both approaches, criticizes their shortcomings as a manifestation of adverse AI, and proposes a way to improve privacy-preserving agents. The article is structured as follows: It dives into personalized agents, specifically their role as privacy-preserving agents and their drawbacks following which, it tackles objective agents in the same regard. Finally, it highlights the proposed approach to improve the existing systems based on our previous findings and other existing research.

## 2. Personalized agent

Some social media platforms such as Facebook offer a native form of personalized settings through which the user gets to set privacy rules for their account, which continue to be applied in the future unless the user makes adjustments to them. These solutions do not often yield much success due to the general sense of *apathy* that a lot of individuals experience online. They do not make the effort not necessarily because they are unaware of the repercussions but because they simply are "privacy fatigued" [9].
A myriad of incidents and breaches are heard of one after the other to the point of users becoming numb to their actions and their repercussions. This is encompassed in the emotional exhaustion and cynicism known as privacy fatigue, which takes a toll on people causing them to reduce or completely shut down their decision-making faculty. Seeing as preference-based settings are not ideal for tackling the issue, there is an ongoing effort towards making the process more tailored yet less engaging on the user part lest they feel burdened. Kurtan *et al.* [10] propose an agent-based approach that leverages the user's self-declared preferences on previous posts to predict the settings for new images. Their proposed system relies on machine learning and learns how to make such recommendations using ideas inspired by information retrieval models. Many privacy scholars examine this as a matter of preferences: how much or little information does this person want to disclose, with whom and for what purpose? [11]. Figure 1 shows the general submodules of a personalized agent, which is adapted from our previous research [12] [13]. When tackling the privacy versus the disclosure appetite dilemma using a personalized approach, our results show that for certain scenarios, up to 93% of Europeans and 86% of North Americans accept the nudges. The system starts by analyzing the user's input for disclosure detection, which means, using Natural Language Understanding models for example, if the post is text-based.

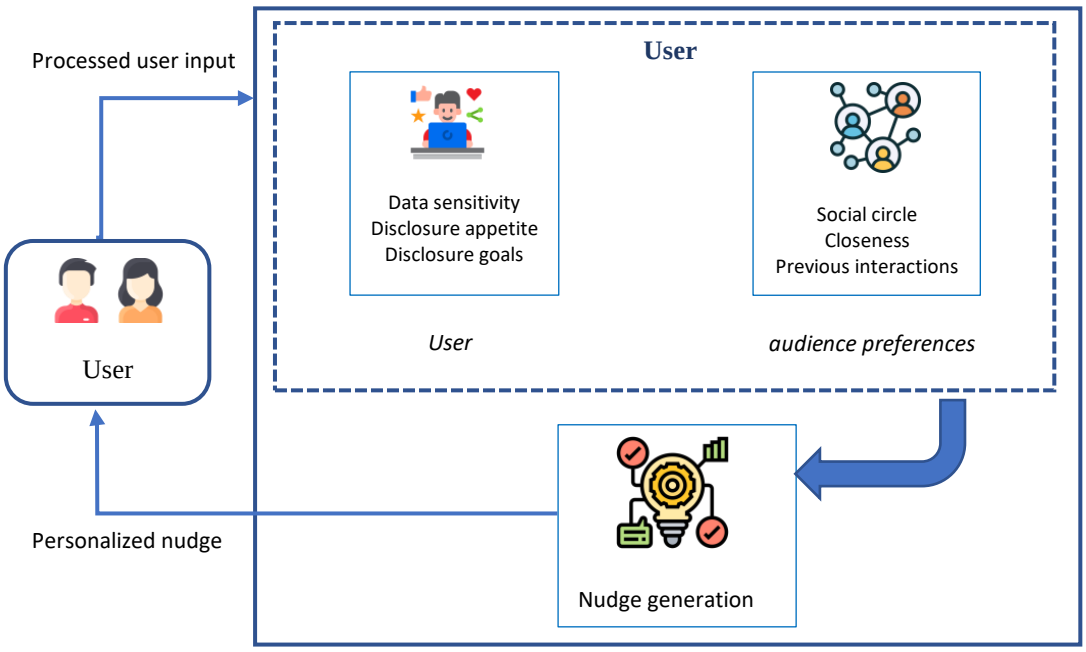


**Figure 1:** Overview of the personalized agent

Following this, the disclosure is compared to the user's self-reported or implicitly inferred preferences (part of the user model) and a nudge is pushed if there is a discrepancy between the two. The preferences include the audience with whom the user likes to share their content, what goals the individual in question aims to achieve through the disclosure, etc.

The classic research of Westin [14], measures privacy with a one-time survey including questions to determine how much accessibility the respondents think is important and how much they value the ownership of data and the control over it [13]. Having these reported user-specific valuations is very crucial to the concept of smart personalized positive behavioural reinforcement. The term "nudge" was popularized by University of Chicago economist Richard Thaler and Harvard Law School professor Cass Sunstein popularized the term "nudge" in 2008 but has gained a new dimension with the recent advances in AI and machine learning. Intelligent personalized tailored nudges are known as "algorithmic nudges" [15] and are capable of being deployed and adapted in real-time. They are very powerful tools for a system meant to target individuals and cater to their perceptions and needs. The user would not be dissatisfied if the system is designed for their specific needs. On the other hand, the concept of algorithmic privacy-preserving nudges immediately poses the question: "what becomes of this if the user does not value their data sensitivity adequately?". In Figure 1, this is represented as part of the "user preferences" but it is worth noting that it is tied to their knowledge and background. This paper does not dive deep into the correlation between parameters like age, gender, occupation, and their impact on self-reported preferences. Instead, it considers that if the user lacks the proper knowledge, regardless of its origin and causes, it causes them to be unable to identify what qualifies as sensitive. They might be driven by their desired goal and disclosure appetite as we call it [13] without realizing the repercussions. This is the first drawback of such a meticulously tailored intelligent agent. From this point onwards in the article, we will use two main scenarios to articulate the arguments:

**Scenario 1**: "Bob and his friends Sam and Alex are about to graduate and will be *looking for employment* soon. Bob often shares social media posts with his friends and enjoys *feeling closer to them* through these interactions. One of which is instigating conversations that end up sparking *controversial debates* between the three of them involving politics, religion, human rights, etc."

**Scenario 2**: "Alice's friends often share the glamorous side of their life such as their luxury high fashion clothes and accessories. She *feels compelled to do the same* and makes sure to post about her extravagant vacations and frequent lavish dining experiences that she splurges most of her income on. She recently decided to *apply for a mortgage* and become a homeowner".

Let us take the example of the second scenario: if asked to answer a questionnaire, Alice who is used to seeing others overshare without any immediate consequences is likely to answer that she is being careful. She perceives the situation to be safe and that her sharing behaviour is not detrimental. She lacks the knowledge to properly identify the issue, especially with regard to her aim to become a homeowner. According to Simon Conn, an overseas mortgage expert with 35 years of experience in the financial services industry, one of the red flags based on which a mortgage application can be rejected is a "boast of excessive lifestyle habit" [16]. Alice, unaware of this, would report that what she is constantly sharing is acceptable and a solely preference-based agent would not disagree nor push nudges to deter her or mitigate her actions. We could say that it even further perpetuates this behaviour and makes Alice more at ease with it since this "protector" allows it. Moving on from this, lacking knowledge is not the only thing that potentially makes a personalized agent underperform or even cause an opposite effect with regards to protecting the user. Another issue arises from the fact that artificially intelligent privacy-preserving interventions that rely on preferences are generally based on the construct of privacy calculus. These approaches assume that as long as the user possesses the required knowledge, their preferences would reflect this, and they will act in a reasonable manner corroborated by the privacy calculus. The issue is that despite many studies such as the one by Fehrenbach *et al.* [17] showing that consumers distinguish between the positive and negative consequences when they determine the value of their data, users still make ill-advised decisions. A compromised knowledge is not the only drive for disclosure. Just because people demonstrate an ability to identify danger in a hypothetical situation does not mean that they would avoid it when it becomes a reality. This is where *cognitive biases* play a major role:

rationality is often overridden by users' immediate gain, appetite for disclosure, and practicality. In fact, it comes second to situational parameters when facing a real scenario and the discrepancy between that knowledge and the action or attitude at that moment is called the *privacy paradox*. "What people decide their data is worth depends critically on the context in which they are asked - specifically, how the problem is framed" [18]. Many individuals who are concerned about their privacy will often voluntarily reveal information to others.

To highlight the fact that this is by no means a laymen issue, an interview with 20 privacy and cybersecurity experts about their views on online privacy is conducted [19]. It revealed that despite their technical knowledge, they end up making the same decisions as non-experts. Even the privacy-conscious might seek instant gratification from disclosure over the long-term loss of assets. The findings of the laboratory study by Ostendorf *et al.* [20] corroborate this by indicating that high self-disclosure via posts is associated with a general tendency to neglect long-term risks. Both internal and external factors can encourage a person to alter their behaviour or feel compelled to make a certain decision. Another example of this can be demonstrated through the *anchoring effect*, which is a form of cognitive bias that causes people to focus on the first available piece of information (the "anchor") given to them when making decisions. Chang *et al.* [11] put this in perspective by asking the users how likely they are to disclose personal information right after seeing examples of increasingly "risky" selfies. Applying this to our first scenario with Bob as the central actor, if he logs into his social media account and the first post he sees shows a couple of his friends having a political debate, that could serve as the anchor and he might proceed to do the same. The second scenario showcases a different type of bias, which is the *bandwagon effect*, a psychological phenomenon in which people do something primarily because other people are doing it. Alice is following a "trend" amongst her friends and ends up adopting a similar oversharing behaviour to them. Other examples of biases are detailed in section 3.

The next point details how tailored agents consider the preferred audience as a personalization parameter. Social circles are represented using structures such as a trust-based graph that accounts for the connection strength between the sharer and the person with whom the information is disclosed [21]. This is determined by characteristics that indicate their level of closeness such as friendship duration, the different social circles they belong to, their previous interactions, etc. The main issue with such a system is that trust is very hard to measure, which makes it unreliable as a way to personalize nudges. Just because Bob has been friends with Philippe on social media for years and they had shared interests, it does not mean that they are very close. They might have never even met in real life and an agent, regardless of how smart it is, would not fully grasp the strength of their friendship. It might then proceed to reassure Bob that it is fine to share the content with Philippe and 200 other people deemed to be trustworthy. Bob might revise this decision if the audience was not so large but in such a situation, he is likely to go along with the convenient nudge thinking it is the best decision for his privacy. Moreover, even if there were a definitive way to calculate the best-fit audience without a shadow of a doubt, another issue can occur. The proposed nudge does not account for the unintended audience that has gained access to the information through a re-share/retweet [22]. The shared information can be accessible to those outside of the imagined audience [23], a concept that can be understood as "the people that the user intended to share with". In other words, there are no guarantees, in this case, that the post will not reach unintended audience and this could be very problematic depending on how sensitive the information is. If Bob is sharing the post with his friends amongst whom John decides to re-share his highly controversial post, it would reach a wider audience including John's friends. This is known as "social contagion", which is the diffusion of information through a network, which has been perceived for a long time to be analogous to the spread of a viral epidemic. When applied to Bob's case (first scenario), his divisive opinions can impact all aspects of his life especially when he is a job seeker. Companies do not want to be associated with contentious issues, especially in today's world.

Finally, we have gone through many pros and cons of using agents that are completely preference-based. The next section dives into the opposite side of the spectrum of privacy-preserving agents and precisely to objective non-user-based assistants. As Acquisti *et al.* [24] explain, there is a need to direct efforts towards developing intelligent agents for the best interest

of individuals beyond preference-based approaches. This is corroborated by our findings [13], which do show a good acceptance rate of the personalized nudges but simultaneously validate the privacy paradox and the biases that users are susceptible to.

## 3. Objective agent

Prior to being used for privacy purposes, objective intelligent agents have been used for bias detection, misinformation identification and deprogramming thanks to their neutral nature [25]. A good argument for their use is that it seems contradictory to counter bias using a biased subjective user-serving agent. First of all, what is an objective agent and how does it differ from its personalized counterpart? The latter relies on the perception of the user, whether that is based on an explicit preference elicitation process (questionnaire for example) or deduced from past posts. However, the former is not centred around the sharer's preferences. It might use objective parameters like the number of friends they have on social media to highlight the reachability of the post but it does not consider the sharer's disclosure goal or sought-after gratification. In this context, we previously proposed Aegis [26] as a means to nudge users towards ethically compliant behaviour on social media.

It is inspired by the consequentialist approach, specifically how the risk and negative outcome weighs on the morality of the action. Another example of approaches that focus on the content rather than the preferences is the work by Battaglia *et al.* [27], which relies on text categorization to calculate a privacy score associated with the current disclosure while drawing inspiration from sentiment analysis. For further explanation, let us envision how an objective agent would act when put in the situation described in the first scenario. Aegis could ask Bob to cease the disclosure because he is involving Alex and Sam and his benefit does not outweigh their loss of privacy. Alice, whose behaviour is detailed in scenario 2, could also benefit from an agent that aims to push her to stop sharing too much about her spending behaviour and so many facets of her lifestyle. The way to achieve this is through incorporating domain knowledge as the drive for nudge rather than the user preference as seen in Figure 2. It is mainly inspired by our Aegis system [26], which bases its interventions on non-user-specific metrics utilizing the existing literature and studies as well as reports by Experian, TransUnion, Atlas VPN, Safety detectives, etc.

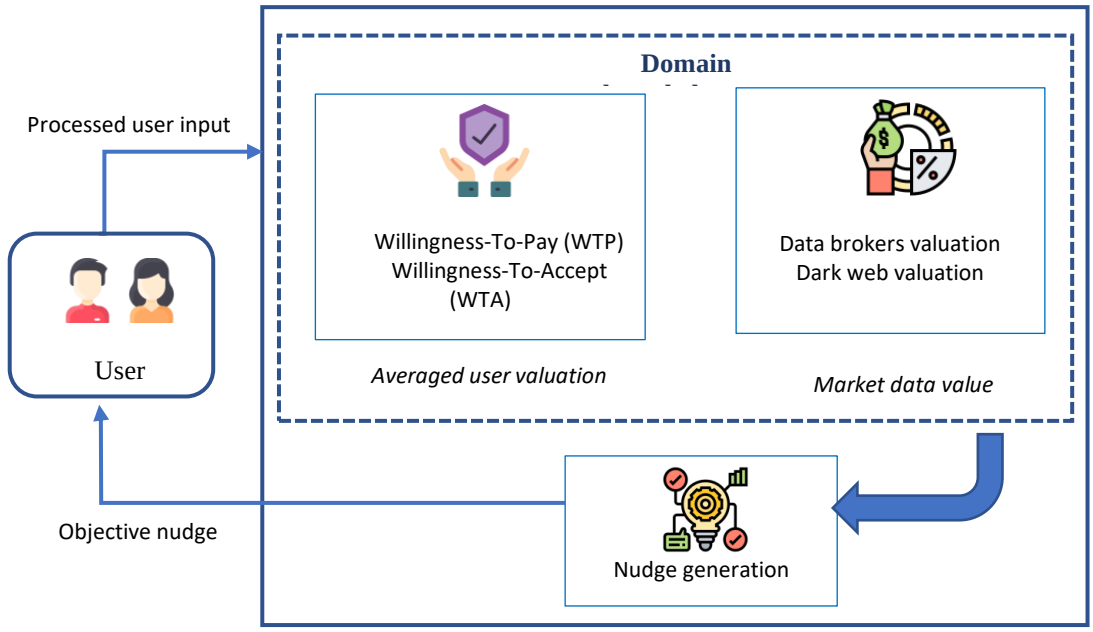


**Figure 2:** Overview of the objective privacy agent

The first main component is the averaged user valuation, which can also be interpreted through the metrics Willingness-To-Pay (WTP) for privacy and Willingness-To-Accept (WTA) payment in exchange for disclosing private information. Previous research has shown that people assign a higher monetary value to privacy in the WTA condition when compared to the WTP condition [28]. The second component is the market data value and is broken down into two subcategories: first, the value of data for companies that acquire it legally and second, its value on the dark web. This is meant to provide a comprehensive perspective of the economic value and the benefit for third parties generated from collecting and mining data from users. Combining the aforementioned components, Aegis reports promising results in estimating the potential loss of privacy due to sharing personal information based on 800 consecutive simulations we ran. This is the foundation of objective non-user-specific nudges.

Going back to the supporting arguments for the objective approach, at face value, it seems to be the most protective measure when it comes to both the individual and their surroundings because it focuses on privacy over appeasing the user. Nevertheless, it should be highlighted that such a system does not force the sharer so regardless of how "good" the nudge is if it is not accepted, it is meaningless. In the worst-case scenario, the user, whether it is Bob (scenario 1) or Alice (scenario 2) would feel too restricted by this system and just deactivate it completely. This is the main weakness of objective agents: they do not account for the human factor such as the disclosure goal in the equation even though it is up to that same human to accept/reject the nudge. One of these well-studied aims or objectives of self-disclosure has been studied long before the advent of social media is developing and maintaining relationships. In 1973, psychologists Irwin Altman and Dalmas Taylor formulated the theory *of social penetration* [29]. It theorizes that the more people disclose things about themselves, the closer they get to those with whom they share said information. This applies to friendships [30] and romantic [31] relationships where this reciprocal act is regarded as necessary to build and maintain interpersonal ties. Going back to the scenario with Bob, his friendship with Sam and Alex deepened the more they bonded over their shared common interests and participated in debates. An objective system that does not consider this and simply nudges him to not disclose this is not likely to be well-received by him.

Aside from the development of interpersonal relationships, people reveal intimate and personal information for a multitude of reasons [32]. One of them is social proof, which is driven by the assumption that the surrounding people possess more knowledge about the current situation. This leads the individual to not only publicly comply but also privately accept the group "knowledge". In the second scenario, Alice is influenced by the numerous people around her who overshare their lifestyle to the point of her ending up conforming to the same actions. In addition to this public compliance, her inner thoughts follow suit and she starts feeling more at ease internally with this behaviour. The objective agent would try to push nudges to Alice but unbeknownst to it, Alice's susceptibility to social compliance is very high and a generic nudge is not likely to be well received.

Another argument for the conception of objective agents is that it eliminates the need for user modelling. The objective agent that we proposed in [26], does not base its data valuation on user preferences, instead, it uses the market, the dark web, and the average user valuation. The first refers to data being sold legally often as part of a collective dataset for marketing purposes. For example, according to a report by Avast [33], the monetary value can be worth more than $240 per year for a bundle of user details and an email address alone can retail for about $89. The second category includes data sold illegally after being hacked or acquired through social engineering. This can be seen in a positive light because it eliminates the need for user-specific data that personalized preference-based approaches use. Based on this, objective agents are innately more privacy-preserving. However, this is not completely true because of two main reasons: the first has already been detailed which is the need to consider the disclosure goal and the second is that the loss of user-specifics turns each sharing instance into an independent event where: the user Bob is sharing a "piece of data A" regardless of his past behaviour. This overlooks the major impact of self-disclosure, which happens over time. Bob can reveal his location at time t1 and that unique instance is harmless but if he does so frequently it can reveal a pattern and he can be easily tracked. An objective non-tailored agent would not be capable of capturing the correlation

between the different disclosures and their future repercussions and this is a crucial shortcoming.

An additional aspect that can reduce the success rate of such agents lies in the fact that interactivity and a seemingly customized experience appeal to users. This is not solely specific to privacy as in multiple fields, leveraging interactive tailored conversational agents shows great promise. For example, in the healthcare domain, notable positive behavioural changes are associated with such technology like smoking cessation [34]. Cybersecurity and privacy-preserving nudges are not exempt from this, but this cannot be achieved through a completely objective agent that pushes one-size-fits-all nudges. Framing directly impacts the decision maker's conception of acts, outcomes, and contingencies associated with a particular choice [35]. Having gone through the positives and negatives of relying solely on personalized or objective agents, the next section explores what we perceive to be the better option.

## 4. Towards enhanced privacy agents: A balance between objectivity and personalization

It is apparent that using a personalized agent or an objective agent on their own does not achieve the best result. While the former pushes nudges that are more likely to be accepted by the sharer since they are based on their preferences, it further perpetuates biases and does not necessarily fulfil its purpose. If Bob is revealing his private data and his preferences corroborate this, the agent is rendered useless. The latter, meaning the objective agent, pushes nudges that guarantee the elimination of disclosure but only if the user accepts them. So, one is too lenient and caters to the user despite their potential ignorance and the other is too idealistic and removes the human element in a process that is highly dependent on the human in question. What we propose is an amalgamation between two concepts that we have previously explored separately [13] [26]: an enhanced privacy agent combining both objective and personalized approaches seen in Figure 3.

When we focused solely on the user preferences as the basis for pushing nudges [13], the system lacked a voice of reason to tune down biases especially when the disclosure concerns other individuals. Hence, when proposing Aegis [26], our aim was to address this by considering multi-party disclosure and prioritizing the domain knowledge over user modelling. Ultimately, we figured out that both systems can complement each other and that balancing the two is needed.

In fact, the objective agent Aegis and the subjective personalized agent can cooperate to mitigate an issue pertaining to "incomplete or asymmetric information" [36]. Such a situation is the norm in the field of privacy and information security and they occur when the defender, in our case either of the protective agents, may not know which vector the adversary will use for the attack. If a malicious party intends to use social engineering techniques on a potential victim, both agents on their own lack the "apparatus" to defend the user. This type of attack usually combines exploiting human weaknesses and biases along with objective domain knowledge such as what pieces of data are most vulnerable and also valuable. Hence, combatting it requires incorporating concepts from both the personalized and the objective agent. Through this, a balanced approach can be achieved, which neither alienates the user nor overlooks their vulnerabilities in favour of their preferences. To ensure that the user is not deprived completely of the gratification they are seeking from the disclosure goal, the mediator agent mitigates the action rather than strictly eliminating it. One way to do this is to go past the binary solutions: either "share" or "do not share" the content. This largely undermines the desire of the decision-maker to share a specific post. Such drive can outweigh consideration for privacy especially when the user is motivated by instant gratification. The proposition by Ben Salem *et al.* [13] achieves this through a trade-off between the user's preferences, motivations and data sensitivity. This is an interpretation of the privacy calculus theory that compares risks and benefits and also serves as a reminder to the user when they are susceptible to the privacy paradox. The paper makes use of the *Item Response Theory* (IRT) and specifically the *Rasch model* to this end. One of the limitations of this work is that the eventual risk due to the actions of users other than the sharer is not accounted for. In other words, the risk considered does not encompass the whole picture including the unintended audience who can view the private data. For this to offer a comprehensive solution, it needs to be combined with the Aegis agent [26]. That would ensure consideration for the "reach" of a specific post beyond the user's settings and preferences. The enhanced nudges consider the importance of joining "*decisional privacy*", which involves the

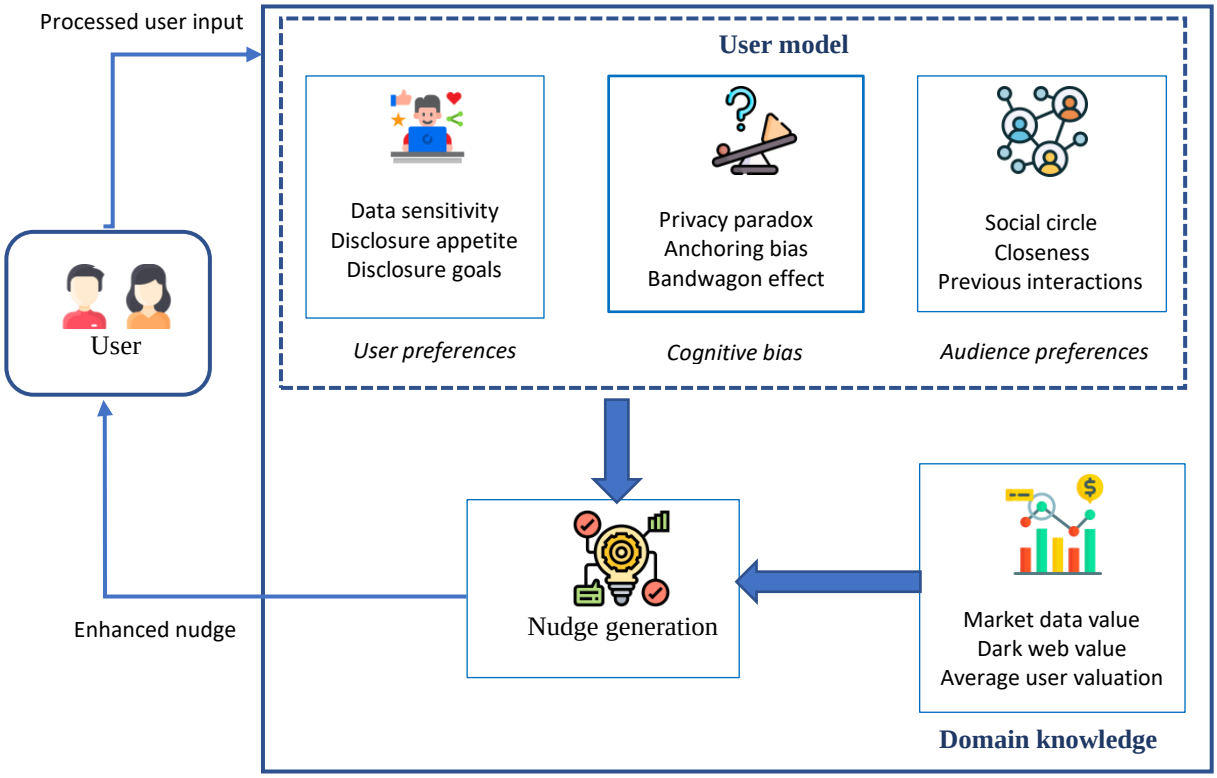


**Figure 3:** Overview of the enhanced hybrid privacy agent

right to independence in making important decisions, and "*informational privacy*", which is founded on the interest in avoiding disclosure of personal matters [37]. There is a fine line between protective educational algorithmic nudging and *hypernudging* leading to compromising the user's critical thinking faculty and their sense of autonomy. Hypernudging [38] is the continuous surveillance and reconfiguration of choice architectures based on large aggregates of personal data which has been facilitated by the progress of machine learning. The user can become complacent and rely on external assistance to push them in the right direction. To avoid this, the solution needs to reinforce long-lasting positive behaviour. Let us consider the scenario where Bob is about to share his controversial opinions. The enhanced hybrid agent knows that Bob is looking for a job thanks to the personalized module and that his disclosure is damaging to that purpose through the knowledge of the objective agent. The latter part could come in the form of statistics or reports such as the one by Jobvite in 2021 [39], which revealed that 30% of recruiters consider sharing political opinions a turn-off when they are reviewing applicants. This percentage goes up to 39% when considering alcohol consumption and 40% when the user has references to marijuana. An enhanced nudge in this situation could be the following: "Did you know that 30% of recruiters consider sharing political opinions a turn-off when they are reviewing applicants? By sharing this post, you would be damaging your potential as a job seeker and that could extend to your friends as well".

Finally, let us circle back to one point that was discussed in Section 3, which is the advantage of objective nudges over their subjective counterparts because they do not require a user model as an input. Now, that this paper suggests combining both approaches, that drawback arises again: How to overcome the fact that user-specific nudges rely on a large amount of personal data and observations of the sharing behaviour? We propose the use of algorithms such as *differential privacy*, which has been widely relied on by personalized recommender systems to protect users' privacy [40]. To the best of our knowledge, this concept has not been adapted to nudges but has the potential to greatly benefit the proposed

enhanced privacy-preserving nudges. If the users are grouped in small communities and noise is added strategically to shield their unique history records and predicted or reported preferences from leakage, then, the major drawback of user modelling can be overcome. However, we acknowledge that this is not a simple matter of applying an existing algorithm since differential privacy is a relatively new strategy. It is not without its challenges like the complications of adding noise to the performance of the predictive model.

To summarize this section, both types of agents have their highlights and negative points. A collaborative effort between the two is needed to balance the need for disclosure with privacy protection.

## 5. Conclusion

Privacy in the age of social media and digital services has become more compromised than ever. The youngest generation Z is the first to grow up without knowing a pre-internet world and to this demographic, the culture of oversharing is all that they have ever known. The normalization of sharing personal intimate thoughts, beliefs, and struggles online make individuals over-disclose without any awareness of the repercussion until it is too late. The digital world does not erase or forget anything that has been shared before. However, to disclose or not to disclose is not a black and white issue. The same user facing the same scenario at different times or contexts, in general, can end up making polar opposite decisions [41] [42].

The goal of objective AI solutions is to offer a voice of reason. They solely aim to preserve the user's privacy and that of other people involved. The subjective approaches are generally more acceptable to the user and consider their disclosure goals. Nevertheless, on their own, both approaches are not enough and can even be detrimental, in which the user-specific subjective approach becomes nothing but an echo of the user's self-declared preferences that can result from a lack of knowledge. Furthermore, personalized assistants can give agency to cognitive biases and make the presence of the privacy paradox even more prominent. Objective agents are too restrictive, and strict and do not offer the user the option to partially quench their appetite for sharing. This paper analyzes both types of privacy-preserving concepts and concludes that one cannot exist without the other and that both are needed to overcome the adverse impact of this technology. We are of the mind that an intelligent context-aware privacy-preserving nudge-based system is needed to remedy the aforementioned shortcomings.

## 6. Acknowledgement

This work is supported in part by Canada's Natural Sciences and Engineering Research Council (NSERC).

## 7. References

[1] Li P-P., Zhong F., "A Study on the Correlation Between Media Usage Frequency and Audiences' Risk Perception, Emotion and Behavior, " Frontiers in Psychology, vol. 12, 2022.

[2] Acquisti A., Adjerid I., Balebako R., Brandimarte L., Cranor L., Komanduri S., Leon P., Sadeh N., Schaub F., Sleeper M., Wang Y., Wilson S., "Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online," ACM Computing Surveys, vol. 50, no. 3, pp. 1-41, 2018.

[3] Jesse M. W., Jannach D., "Digital Nudging with Recommender Systems: Survey and Future Directions," Computers in Human Behavior Reports, vol. 3, 2021.

[4] Halpern D., Inside the Nudge Unit: How small changes can make a big difference, 2016.

[5] Kankane S., DiRusso C., Buckley C., "Can We Nudge Users Toward Better Password Management?: An Initial Study," in CHI Conference on Human Factors in Computing Systems, pp. 1-6, 2018.

[6] Peer E., Egelman S., Harbach M., Malkin N., Marthur A., Frik A., "Nudge me right: Personalizing online security nudges to people's decision-making styles," Computers in Human Behavior, vol. 109, 2020.

[7] Díaz Ferreyra N.E., Kroll T., Aïmeur E., Stieglitz S., Heisel M., "Preventative Nudges: Introducing Risk Cues for Supporting Online Self-Disclosure Decisions," Information, vol. 11, no. 8, pp. 399, 2020.

[8] Kroll T., Stieglitz S., "Digital nudging and privacy: improving decisions about self-

disclosure in social networks," Behaviour & Information Technology, vol. 40, no. 1, pp. 1-19, 2019.
[9] Shao H., Li X., Wang G., "Are You Tired? I am: Trying to Understand Privacy Fatigue of Social Media Users," in CHI Conference on Human Factors in Computing Systems, pp. 1-7, 2022.
[10] Kurtan A., Yolum P., "Assisting humans in privacy management: an agent based," Autonomous Agents and Multi-Agent Systems, vol. 35, no. 1, 2021.
[11] Chang D., Krupka E., Adar E., Acquisti A., " Engineering Information Disclosure: Norm Shaping Designs," in CHI Conference on Human Factors in Computing Systems, pp. 587-597, 2016.
[12] Ben Salem R., Aïmeur E., Hage H., "The Privacy versus Disclosure Appetite Dilemma: Mitigation by Recommendation," in Workshop on Online Misinformation- and Harm-Aware Recommender Systems, ACM Recommender Systems Conference (RecSys), 2021.
[13] Ben Salem R., Aïmeur E., Hage H., "A Nudge-based Recommender System Towards Responsible Online Socializing," in Workshop on Online Misinformation- and Harm-Aware Recommender Systems, ACM Recommender Systems Conference (RecSys), 2020.
[14] Westin A., "Privacy and Freedom," in Athenaeum, 1970.
[15] M. Möhlmann, "Algorithmic Nudges Don't Have to Be Unethical," Harvard Business Review: Analytics And Data Science, 2021.
[16] Conn S., "Social media impact on credit rating and mortage applications," Simon Conn oversees property and financial specialist, 24 11 2020. [Online]. Available: https://www.simonconn.com/blog/social-media-impact-credit-rating-overseas-mortgage-applications/. [Accessed 13 05 2022].
[17] Fehrenbach D., Herrando C., "The effect of customer-perceived value when paying for a product with personal data: A real-life experimental study," Journal of Business Research, vol. 137, pp. 222-232, 2021.
[18] John L., Acquisti A., Loewenstein G., "Strangers on a plane: context dependent willingness to divulge sensitive information," Journal of Consumer Research, vol. 37, pp. 858–873, 2011.
[19] Barth S., Jong M., Junger M., "Lost in privacy? Online privacy from a cybersecurity expert perspective" Telematics and Informatics, vol. 68, pp. 101782, 2022.
[20] Ostendorf S., Müller S.M., Brand M., "Neglecting Long-Term Risks: Self-Disclosure on Social Media and Its Relation to Individual Decision-Making Tendencies and Problematic Social-Networks-Use," Frontiers in Psychology, vol. 11, 2020.
[21] Voloch N., Gal-Oz N., Gudes E., "A Trust based Privacy Providing Model for Online Social Networks," Online Social Networks and Media. vol. 24, pp. 100138, 2021.
[22] Alemany J., Del Val E., Alberola J., García-Fornes A., "Estimation of privacy risk through centrality metrics," Future Generation Computer Systems, vol. 82, pp. 63-76, 2018.
[23] Litt E., Hargittai E., "The Imagined Audience on Social Network Sites," Social Media + Society, vol. 2, no. 1, 2016.
[24] Acquisti A., Brandimarte L., Hancock J., "How privacy's past may shape its future," Science, vol. 375, no. 6578, pp. 270-272, 2021.
[25] Chen W., Pacheco D., Yang K., "Neutral bots probe political bias on social media", Nature communications, 12, 5580, 2021.
[26] Ben Salem R., Aïmeur E., Hage H., "Aegis: An Agent for Multi-party Privacy Preservation", AAAI/ACM Conference on AI, Ethics, and Society, 2022.
[27] Battaglia E., Bioglio L., Pensa R.G., "Towards Content Sensitivity Analysis," Advances in Intelligent Data Analysis XVIII, 2020.
[28] Acquisti A., John L., Loewenstein G., "What Is Privacy Worth?," The Journal of Legal Studies, vol. 42, no. 2, pp. 249–274, 2013.
[29] Taylor A., Altman I., "Social penetration: The development of interpersonal relationships," New York: Holt, 1973.
[30] Jourard S., "Self-disclosure: An experimental analysis of the transparent self," Florida, Oxford, England: John Wiley, 1971.
[31] Kil H., Allen M-P., Taing J., Mageau G.A., "Autonomy support in disclosure and privacy maintenance regulation within romantic relationships," Personal Relationships, vol. 29, no. 2, pp. 305-331, 2022.
[32] Aïmeur E., Díaz Ferreyra N.E., Hage H., "Manipulation and Malicious Personalization: Exploring the Self-

Disclosure Biases Exploited by Deceptive Attackers on Social Media," no. https://www.frontiersin.org/articles/10.3389/frai.2019.00026/full.
[33] Latto N., "Data Brokers: Everything You Need to Know," Avast, 25 11 2021. [Online]. Available: https://www.avast.com/c-data-brokers. [Accessed 21 02 2022].
[34] Calvaresi D., Calbimonte J., Dubosson F., Najjar A., Schumacher M., "Social Network Chatbots for Smoking Cessation: Agent," in IEEE/WIC/ACM International Conference on Web Intelligence (WI), 2019.
[35] Tversky A., Kahneman D., "The framing of decisions and the psychology of choice," in Environmental Impact Assessment, Technology Assessment, and Risk Analysis, pp. 107-129, 1985.
[36] Acquisti A., Adjerid I., Balebako R., Brandimarte L., Cranor L.F., Komanduri S., Leon P.G, Sadeh N., Schaub F., Sleeper M., Wang Y., Wilson S., " Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online," in Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online, vol. 50, no. 3, pp. 1-41, 2018.
[37] Koops B.J., Newell B., Timan T., Skorvánek I., Chokrevski T., Galič M., "A Typology of Privacy", University of Pennsylvania Journal of International Law, vol. 38, no. 2, 2017.
[38] Yeung K., "'Hypernudge': Big Data as a mode of regulation by design," Information, Communication & Society, vol. 20, no. 1, pp. 118-136, 2017.
[39] Jobvite, "Agility: The Essential Ingredient for Recruiting Success," Available: https://www.jobvite.com/wp-content/uploads/2021/09/Jobvite-RecruiterNation-Report-WEB-2.pdf [Accessed 13 05 2022].
[40] Wang H., He K., Niu B., Yin L., Li F., "Achieving Privacy-Preserving Group Recommendation with Local Differential Privacy and Random Transmission," Wireless Communications and Mobile Computing, 2020.
[41] Parrey B. Skinner G., "A literature review on effects of time pressure on decision making in a cyber security context," Journal of Physics Conference Series, vol. 1195, 2019.
[42] Chowdhury N. H., Adam M., Skinner G., "The impact of time pressure on cybersecurity behaviour: a systematic literature review," Behaviour & Information Technology, vol. 38, no. 12, pp. 1290-1308, 2019.