

Mitigating Digital Mindlessness

Sushmita Khan¹, Mehtab Iqbal¹, Nushrat Humaira¹, Nina Hubig¹ and Bart Knijnenburg¹

¹*School of Computing, Clemson University, 821 McMillan Rd, Clemson, SC, 29631, USA*

Abstract

Mindfulness is a well-studied phenomenon and the use of technology to promote mindfulness are quickly gaining popularity. However, in the digital space mindfulness is often overlooked. Instead, there is a rise in deceptive and exploitative techniques in artificial intelligence (AI) that encourage and exploit users' *mindless* behaviors. To combat digital mindlessness and the effects of these hazardous AI, this position paper explores the literature in an attempt to define the concept of mindlessness and find ways to reduce it (i.e., to promote more mindful behavior). Particularly, mindless-making effects of AI is explored from four thematic viewpoints 1) *Filter Bubbles*, 2) *Dark Patterns*, 3) *Fake news and Misinformation*, and 4) *Deepfakes*. We end the paper by identifying potential lines of research to aid in the reduction of mindlessness.

Keywords

Digital Mindfulness, Mindlessness, Filter Bubble, Dark Patterns, Misinformation, Deep Fakes

1. Introduction

There has been an increase in research in Human Computer Interaction (HCI), Artificial Intelligence (AI), and related computational fields on the use of technology to promote mindful behavior [1]. There are many definitions of mindfulness that are used in literature but for our work we follow the definition of Terzimehić et al. who proposed four distinct areas, and seven “lines” of research in their survey [1]. The four categories are (1) *Role* focused on achieving goal through mindfulness, (2) *Practice* involved both direct activities like meditation and implicit ones like mitigating distraction (3) *Longevity* involved studies with either short-term or long-term goals of mindfulness, and finally (4) *Co-Aspects* was mostly about reflection but often used interchangeably with the term mindfulness itself. The four categories were distilled to resolve the, often non-complementary, set of definitions that exist in the current literature, and thus provide goal based definitions. Similarly, the seven research themes they uncovered were (A) *Meditation Practice*, (B) *Therapy*, (C) *Reflection & Knowledge Gain*, (D) *Mindfulness in Daily Life*, (E) *Mindfulness in Interaction*, (F) *Performance Enhancement*, and (G) *Meta-Level Research*. These proposed area of research are situated based on the four distinct goal based definitions.

Most mindfulness research leverages digital technologies to improve people's mindful practices in the real world. Although increasingly relevant parts of our daily lives happen online, there is a dearth of research on mindfulness when using (AI) technology or during interactions

AIofAI'21: 1st Workshop on Adverse Impacts and Collateral Effects of Artificial Intelligence Technologies, Montreal, CA

✉ sushmik@clemson.edu (S. Khan); mehtabi@clemson.edu (M. Iqbal); nhumair@clemson.edu (N. Humaira); nhubig@clemson.edu (N. Hubig); bartk@clemson.edu (B. Knijnenburg)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

in the digital space. Additionally, lot of automation, and AI based interaction often operate to reduce the cognitive demand of everyday task. In this position paper, we argue that it might be more beneficial to approach mindfulness in the digital realm from the other end—*how can we mitigate digital mindlessness?*

Indeed, the algorithms that govern our online interactions (i.e., by providing personalization, recommendations, or otherwise aiding in online decision making) seem to be responsible for creating *filter bubbles* [2], increasing choice overload [3], and inducing distractions [4]. While seemingly benign, filter bubbles can cause users to get “trapped” by their short-term goals [5, 6], dopamine system exploits and even help propagate *misinformation* [7]. Although regulations like GDPR [8] give users some control over the impact of algorithmic tracking, and their personal data the use of *dark patterns* [9], all but renders this control ineffective by exploiting users’ tendency to blindly accept such notices. Furthermore, beyond these existing algorithmic practices, there is a recent rise in the use of algorithms to create *deepfakes*, which can be used to manipulate or falsify text, images and video in a way that is indistinguishable from the genuine to the inattentive eye.

While mindfulness is multifaceted, we propose a definition for *digital mindlessness* focused on user autonomy and self-actualization [10] that is centered on the loss of agency that occurs due to AI based automation:

Definition. Mindlessness is the vulnerable state of heuristic decision making which leaves users susceptible to unintentional or malicious deception (e.g. fake news, deep fakes) and misdirection (e.g., filter bubbles, dark patterns).

2. Elements of Mindlessness

In this section we discuss four elements relevant to mindlessness in terms of intent, effect, and ingenuity. Filter bubbles are usually not caused by malicious intent but rather an effect of over-personalization. Dark patterns are deliberate actions to exploit others, but vigilance and best practice can mitigate its effects. On the other hand, deepfakes are often indiscernible without contextual knowledge, akin to fake news and misinformation, whose propagation isn’t always voluntary. To illustrate these effects further, we focus on how these mechanisms exploit online users’ lack of awareness to warp their perception of reality.

2.1. Filter Bubbles

Personalized systems are ubiquitous, and they power digital platforms that enable socializing (social media sites), searching for information (Google, Baidu), news consumption, surveillance (employee monitoring), filtering (spam), forecasting (flu trends, weather, stock market), recommendation scoring (credit score, social score), content production (algorithmic journalism), advertisement, etc. [11, 12]. Due to the user-centered nature of these personalized systems, the information they present is deliberately curated to fit users’ preferences. It is often claimed that users benefit from this personalized presentation: personalization leads to increased individualization, commercialization, and deterritorialization [11]. On the other hand, these black-box

AI systems are also known to decrease transparency, controllability, and predictability, and to promote existing inequalities [11]. This phenomenon, known as *the Filter Bubble* [2] has a far-reaching influence on our online lives.

Filter Bubbles [13] exist because online systems use user preference data to present personalized information to users. While this ensures interaction and consumption of content that users are interested in, it also traps them in a bubble of their preferences, isolating them from information that challenges their standing. Such filter bubbles are further reinforced as users continue to interact with their preferred topics in their bubbles, thereby reaffirming their already established preferences [13], ultimately creating an “echo-chamber” [10]. This isolation leads to decreased awareness of reality through a lack of exposure to otherwise important issues and different points of view. This distortion of reality is detrimental as it affects the decision making process of our society as a whole, which in turn impacts public policy [14].

Although AI-based personalization is commonplace in online interaction, research shows that users are only minimally aware of filter bubbles. For instance, [15] find that 61% of the Norwegian population has no or low algorithmic awareness. Given the widespread algorithmic infrastructure that amplifies existing patterns, this lack of algorithmic awareness arguably exacerbates the problem by weakening informed participation, increasing inequalities, deepening the digital divide, and reducing the quality of the online lived experience [16]. Increased awareness through algorithmic literacy [15] coupled with exposing users to a broader spectrum of topics is perhaps a feasible solution to the filtering problem [10].

2.2. Dark Patterns

Recently introduced privacy protections such as the GDPR [8] presumably give end-users more control over the extent to which AI systems govern their daily digital lives. In response, Web site owners have started to employ (or increase their use of) *dark patterns* to continue their current data collection and personalization practices [17]. Dark patterns are the “evil twin” of user interface (UI) and experience (UX) design. UI and UX design leverage knowledge of human behavior to provide affordances that make a system easier to understand and more accessible [18, 9]. This knowledge of human behavior can however also be used to misdirect users towards goals that benefit other stakeholders, often to the detriment of the user [19]. In AI-based systems, dark patterns tend to be used to coerce inattentive users to opt in to the algorithmic personalization practices.

There has been an increase in UI designs with dark patterns embedded in them. For example a recent study analyzing 240 mobile apps found 95% of the apps to be using some sort of dark patterns [20]. Nagging, obstruction, sneaking, interface interference, and forced action are prominent dark patterns that serve as strategic motivators that can nudge inattentive users into accepting intrusive tracking practices. Not all dark patterns are developed with malicious intent [19], but all have ethical implications nonetheless, as they intentionally confuse and manipulate users [21].

2.3. Fake news and Misinformation

Misinformation is false or inaccurate information (e.g., rumors, urban legends, factoids) that covers topics similar to those in the mainstream news media, but that is embedded with ideological messages to target a subset of the population with a particular political leaning [22]. Also known under the popular moniker “fake news”, misinformation is propagated both intentionally and unintentionally—often with the aid of social media platforms and exacerbated by the AI algorithms underlying the news feed of these platforms, which tend to emphasize sensational “click bait” over accurate information [22]. Due to the wide reach and the potential damages it can cause, misinformation is considered a critical issue [7].

Misinformation tends to be planted within news texts or images and circulated through social media [7]. Considering how 68% of US adults use Facebook and 73% use YouTube for daily news consumption according to social media use survey done in 2018, the influence of misinformation is arguably far reaching. Furthermore, most social media users—and in particular younger adults—are vulnerable to misinformation because they are either unable to identify misinformation, influenced to accept information that is shared by their social group, or likely to conform to confirmation bias (i.e., having a strong tendency to accept information that confirms their existing views). As mentioned in Section 2.1, these factors are the ideal circumstances for filter bubbles to form, so that the misinformation is further reinforced by algorithmically controlled content delivery strategies that tend to expose users to content they are most likely to interact with [22].

Misinformation is a regular part of our digital society and it affects our decision making process [14, 7] by influencing the information we consume. To mitigate the challenges of misinformation, research focuses on the detection of misinformative content, the dynamics of the spread of misinformation, content validation, and misinformation management [14, 7]. Despite efforts to contain misinformation, it continues to adversely affect the lives of people who consume online news and information. Many of the challenges posed by misinformation are similar to those posed by filter bubbles and thus it is also likely that more awareness of the progeny of information [13], and the introduction of information curated from outside the user’s filter bubble [10] may offer a solution to the problem of misinformation.

2.4. Deepfakes

The term *deepfakes* describes the use of deep learning to generate fake content [23]. While not all applications of deepfakes are malicious, it is very easy to use deepfakes in deceptive ways, since deepfakes are often difficult to detect [24]. In this position paper we use a broad definition of deepfakes that includes any mechanism that uses deep generative models to create and present information that is not true. For example, text generated through GPT-3 [25] is considered a deepfake, as are any of the image- or video-based deepfakes [26].

Deepfake research is mostly focused on their generation and detection—a never-ending “arms race” of similar algorithmic strategies on either side of the battle [23]. Detection alone is not enough to counter the possible adverse effects of deepfakes, though, because unsuspecting users may not think to challenge the veracity of the faked content. In fact, there are currently no perfect solutions to counter the deepfake problem and it is likely that human society will

eventually have to learn to live with deepfakes by becoming more mindful about them [27]. This can be achieved by educating the user about deepfakes and by training AI researchers and developers about the ethical implications of their work [24].

3. Mindlessness vs Mindfulness

Human beings often resort to simple heuristics when making complex decisions. These heuristic decisions are usually riddled with *cognitive biases* [28] that may be exploited or exacerbated by unintentional or malicious AI systems that deceive or misdirect. As outlined in this paper, dark patterns can cause users to opt in to filter bubbles; confirmation biases can turn these filter bubbles into echo chambers, which in turn increase the malicious impact of fake news and deepfakes as they spread across the Internet.

Research shows that heuristic thinking can be overridden by conscious reasoning [29]. While there already exist mindfulness techniques that teach this active reasoning process, they are often applied to real-world decision-making [1]. With the widespread use of dark patterns, the pervasiveness of algorithms creating filter bubbles that spread misinformation, and the increasing rise of deepfakes, creating mindful awareness is exponentially more challenging in the digital realm.

As such, we argue that it is important to understand the limitations of mindfulness in the context of the current digital landscape. Indeed, while mindlessness is the process of users being unaware, we argue that the onus of this lack of awareness is not entirely on the user. Thus, we argue that a research agenda for *digital mindfulness* should enable the design of systems and interfaces that allow users to be more cognizant of how and where they place their trust.

4. Conclusion

In this position paper we argue that understanding the adverse effects of AI algorithms to help overcome digital mindlessness as a way of enabling digital mindfulness. To give impetus to the digital mindfulness research agenda, we plan to conduct a literature review to elicit how reducing algorithmic influence on mindlessness may encourage mindfulness when interacting with technology. Our review will cover the four themes outlined above: *filter bubbles*, 2) *dark patterns*, 3) *fake news and misinformation*, and 4) *deepfakes*. We aim to not only cover the research on these problems themselves, but also the research into solutions to these problems.

While the four thematic problems outlined in this position paper have mindlessness in common, they differ in terms of adversarial intentions and user involvement. Specifically, *filter bubbles* are not necessarily created by an intention to mislead, whereas in the case of *deepfakes* the deception is intentional. Similarly, while the use of *dark patterns* relies on users dissociating from their content, *misinformation* actually relies on the user actively buying into the embedded ideological message. From the perspective of mindlessness, this means that *filter bubbles* cause mindlessness due to over reliance on heuristics, *dark patterns* rely on mindlessness, *misinformation* depends on not only mindlessness but on the rejection of alternate ideas due to cognitive biases, and finally, *deepfakes* expect users to exhibit a strong confirmation bias.

Despite these differences, we conjecture that due to the underlying cause of these problems—the inattentiveness of the end-user—the proposed solutions to digital mindlessness are likely to have a common intent and perhaps even share similar strategies. It is our intention to leverage this commonality, so that more comprehensive solutions can be proposed.

References

- [1] N. Terzimehić, R. Häuslschmid, H. Hussmann, M. Schraefel, A review & analysis of mindfulness research in hci: Framing current lines of research and future opportunities, in: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.
- [2] E. Pariser, *The filter bubble: How the new personalized web is changing what we read and how we think*, Penguin, 2011.
- [3] D. Bollen, B. P. Knijnenburg, M. C. Willemsen, M. Graus, Understanding choice overload in recommender systems, in: *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 63–70.
- [4] U. Lyngs, K. Lukoff, P. Slovak, W. Seymour, H. Webb, M. Jirotko, J. Zhao, M. Van Kleek, N. Shadbolt, 'i just want to hack myself to not get distracted' evaluating design interventions for self-control on facebook, in: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–15.
- [5] M. D. Ekstrand, M. C. Willemsen, Behaviorism is not enough: better recommendations through listening to users, in: *Proceedings of the 10th ACM conference on recommender systems*, 2016, pp. 221–224.
- [6] N. Seaver, Captivating algorithms: Recommender systems as traps, *Journal of Material Culture* 24 (2019) 421–436.
- [7] M. Almaliki, Online misinformation spread: A systematic literature map, in: *Proceedings of the 2019 3rd International Conference on Information System and Data Mining*, 2019, pp. 171–178.
- [8] G. D. P. Regulation, Regulation eu 2016/679 of the european parliament and of the council of 27 april 2016, *Official Journal of the European Union* (2016).
- [9] S. Greenberg, S. Boring, J. Vermeulen, J. Dostal, Dark patterns in proxemic interactions: a critical perspective, in: *Proceedings of the 2014 conference on Designing interactive systems*, 2014, pp. 523–532.
- [10] B. P. Knijnenburg, S. Sivakumar, D. Wilkinson, Recommender systems for self-actualization, in: *Proceedings of the 10th acm conference on recommender systems*, 2016, pp. 11–14.
- [11] N. Just, M. Latzer, Governance by algorithms: reality construction by algorithmic selection on the internet, *Media, culture & society* 39 (2017) 238–258.
- [12] A. Koenig, The algorithms know me and i know them: Using student journals to uncover algorithmic literacy awareness, *Computers and Composition* 58 (2020) 102611.
- [13] S. Nagulendra, J. Vassileva, Understanding and controlling the filter bubble through interactive visualization: a user study, in: *Proceedings of the 25th ACM conference on Hypertext and social media*, 2014, pp. 107–115.

- [14] M. Fernandez, H. Alani, Online misinformation: Challenges and future directions, in: Companion Proceedings of the The Web Conference 2018, 2018, pp. 595–602.
- [15] A.-B. Gran, P. Booth, T. Bucher, To be or not to be algorithm aware: a question of a new digital divide?, *Information, Communication & Society* (2020) 1–18.
- [16] C. O’neil, *Weapons of math destruction: How big data increases inequality and threatens democracy*, Crown, 2016.
- [17] P. Leon, B. Ur, R. Shay, Y. Wang, R. Balebako, L. Cranor, Why johnny can’t opt out: a usability evaluation of tools to limit online behavioral advertising, in: Proceedings of the SIGCHI conference on human factors in computing systems, 2012, pp. 589–598.
- [18] D. Norman, *The design of everyday things: Revised and expanded edition*, Basic books, 2013.
- [19] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, A. L. Toombs, The dark (patterns) side of ux design, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1–14.
- [20] L. Di Geronimo, L. Braz, E. Fregnan, F. Palomba, A. Bacchelli, Ui dark patterns and where to find them: a study on mobile applications and user perception, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–14.
- [21] J. Luguri, L. J. Strahilevitz, Shining a light on dark patterns, *Journal of Legal Analysis* 13 (2021) 43–109.
- [22] A. Karduni, I. Cho, R. Wesslen, S. Santhanam, S. Volkova, D. L. Arendt, S. Shaikh, W. Dou, Vulnerable to misinformation? verifi!, in: Proceedings of the 24th International Conference on Intelligent User Interfaces, 2019, pp. 312–323.
- [23] Y. Mirsky, W. Lee, The creation and detection of deepfakes: A survey, *ACM Computing Surveys (CSUR)* 54 (2021) 1–41.
- [24] D. G. Johnson, N. Diakopoulos, What to do about deepfakes, *Communications of the ACM* 64 (2021) 33–35.
- [25] R. Dale, Gpt-3: What’s it good for?, *Natural Language Engineering* 27 (2021) 113–118.
- [26] B. Chesney, D. Citron, Deep fakes: A looming challenge for privacy, democracy, and national security, *Calif. L. Rev.* 107 (2019) 1753.
- [27] S. Dack, Deep fakes, fake news, and what comes next, *The Henry M. Jackson School of International Studies News* (2019).
- [28] A. Tversky, D. Kahneman, Judgment under uncertainty: Heuristics and biases, *science* 185 (1974) 1124–1131.
- [29] J. S. B. Evans, D. E. Over, Heuristic thinking and human intelligence: a commentary on marewski, gaissmaier and gigerenzer, *Cognitive processing* 11 (2010) 171–175.