OVERCOMING INFORMATIONAL VULNERABILITIES BY MEANS OF AI-DRIVEN METHODS

-A REVIEW OF DESIGN TECHNIQUES-

Abstract— **This study aims to overcome the gap between HCI practitioners and academicians by providing a review of existing design methods and suggests an AI-driven model development to help recognize dark design patterns. Given design privacy challenges amidst big data, it suggests a design model to recognize and analyze usage patterns including image and text variables. This study also suggests a taxonomy while applying learnable prompts in a continuous space, achieving excellent performance on transfer learning.**

The study concludes that by designing choice mechanisms that are meaningful and aware of data subjects' informational vulnerabilities, users can feel more empowered.

Keywords—AI, big data, ML

I. INTRODUCTION

Information often comes in patterns whether we realize it or not. In order for it to be persuasive, designers and engineers often make use of patterns including different semantics, such as anti patterns [14] and dark patterns [11].

This study aims to provide a review of existing design methods and suggests an AI-driven model development to help recognize dark design patterns. It starts with a review of existing studies to explore design patterns. After an overview of the related conceptual model, it suggests the use of a design model to analyze usage patterns including image and text variables. This study also suggests a taxonomy while applying learnable prompts in a continuous space, achieving excellent performance on transfer learning.

II. REVIEW OF EXISTING STUDIES

The idea of a pattern is to capture an instance of a problem and a corresponding solution, abstract it from a specific use case, and shape it in a more generic way, so that it can be applied and re-used in various matching scenarios. These scholars also recognize the possibility that "dark" versions of these design outcomes may exist, and propose a series of design principles to guard against dubious behaviors such as dual-privacy, disclosure, accuracy, and the "golden" principle [12].

The term dark pattern was first used by Brignull, who collected malicious user interface patterns [11] for better awareness. A UI dark pattern tricks users into performing unintended and unwanted actions, based on a misleading interface design.

A dark pattern consists of user interface design choices that manipulate the data subject's decision-making process in a way detrimental to his or her privacy and beneficial to the service provider. An important part of the study of a dark pattern is understanding the cognitive biases they exploit.

A phenomenon known as "anti-patterns" (c.f., [43])—many of these dark patterns result from explicit, purposeful design intentions. This typology mixes context, strategy, and outcome, making comparison among patterns difficult.

Anti-patterns often target solutions that may seem obvious to the system developer at a first glance, but include a number of less obvious negative implications and consequences. It consists of the following patterns:

- *Nagging:* Nagging often manifests as a repeated intrusion during normal interaction, where the user's desired task is interrupted one or more times by other tasks not directly related to the one the user is focusing on.

- *Obstruction:* This is referred as impeding a task flow, making an interaction more difficult than it inherently needs to be with the intent to dissuade an action.

- *Intermediate Currency:* This is another subtype of obstruction where users spend real money to purchase a virtual currency which is then spent on a good or service.

- *Sneaking:* This can be defined as an attempt to hide, disguise, or delay the divulging of information that has relevance to the user.

- *Interface Interference:* This refers to any manipulation of the user interface that privileges specific actions over others, thereby

confusing the user or limiting discoverability of important action possibilities (c.f., false or hidden affordances [32]).

- *Hidden Information:* This include information as options or actions relevant to the user but not made immediately or readily accessible. Hidden information may manifest as options or content hidden in fine print, discolored text, or a product's terms and conditions statement.

- *Preselection:* Preselection usually manifests as a default choice that the shareholder of the product wishes the user to choose; however, this choice is often against the user's interests or may provide unintended consequences. The user is more likely to agree to the default option if they believe the product has their best interests in mind.

- *Aesthetic Manipulation:* Aesthetic manipulation is any manipulation of the user interface that deals more directly with form than function. This includes design choices that focus the user's attention on one thing to distract them from or convince them of something else.

- *Forced Action:* This refers to any situation in which users are required to perform a specific action to access (or continue to access) specific functionality.

Given the use of these different pattern types, AI algorithms rely on unintuitive criteria and non-traditional data which mainly rely on the following conditions when it comes to decision-making:

1. *Opacity:* Individuals may lack information about the purpose of an AI system, the scope and source(s) of data under consideration, the content of the data, the criteria used to make decisions (e.g., thresholds for classifying high- and low-risk cases), and other aspects.

2. *Vagueness:* Similarly to opacity, vagueness describes a condition of profiling or decision-making processes in which the subject receives inadequate information to make informed choices.

3. *Instability:* Many AI systems are not stable, meaning they change over time or produce erroneous or unpredictable behaviors (i.e., edge cases).

4. *Involuntariness and invisibility:* Many data points used by AI profiling and decision-making systems are based on involuntary and invisible digital and physiological behaviors that are not self-evidently meaningful.

5. *Lack of social concept:* The assembly of pixels in a picture, or clicking patterns are examples that do not have a related social concept; society does not currently distinguish between people or groups in these terms or find these characteristics socially salient.

Choice is the main voice of the data subject in the daily interaction with the controller; without meaningful choice, the data subject has a weak presence and existing informational vulnerabilities are exacerbated. Even if there are accessible privacy notices, absent meaningful privacy choices, data subjects will be unable to exercise their autonomy.

To provide the user with different choices, numerous framings of design and values have been explored in the HCI community and beyond in the last two decades (e.g., [7, 25, 27, 28, 30, 60, 62, 64]). Below is a short description of these models.

Value-Sensitive Methods

Value Sensitive Design (VSD) has been one of the most comprehensive frameworks developed to address the question of values in design, described by its creators as "a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process" [29].

Critical and Reflective Design

Critical design builds upon traditional design practices, but rather than resulting in artifacts that affirm current societal norms, the designer creates artifacts or experiences that allow key societal norms and values to be openly interpreted and questioned [21]. Bardzell et al. [10, 9] have previously proposed an approach to analyzing critical designs, building upon both a corpus of exemplars [23] and patterns of humanistic interpretation [8] to foreground critical dimensions of these artifacts.

Persuasive Design

Design is inherently a persuasive act [54, 57, 67], where the designer creates intentional change in the world that either directly or indirectly induces behavioral or social change. Fogg [27] views persuasive technology as "[designing for] behavior as something we cause to occur [... and/or] preventing a target behavior from happening." This shaping of behavior is proposed to be accomplished through seven persuasive strategies: reduction, tunneling, tailoring, suggestion, self-monitoring, surveillance, and conditioning [26].

To verify the effectiveness of data filters, generative models with the same hyperparameters can be trained on both unfiltered data on the dataset after filtering. Generative models attempt to match the distribution of their training data, including any biases therein. As a result, filtering the training data has the potential to create or amplify biases in downstream models.

Since training data shapes the capabilities of any learned model, data filtering is a powerful tool for limiting undesirable model capabilities. This approach can be implemented in two categories—images depicting inappropriate content—by using classifiers to filter images in these categories out of the dataset before training and train these image classifiers in-house.

According to Hoepman [24], a privacy design strategy is on a more general level than a privacy pattern and "describes a fundamental approach to achieve a certain design goal. It has certain properties that allow it to be distinguished from other (fundamental) approaches that achieve the same goal." Hoepman [24] defines the following eight privacy design strategies:

- *Minimize:* Data minimization is a strategy which insists that the amount of personal information that is processed should be minimal. Data that is no needed for the original purpose should not be collected.

- *Hide:* Hide takes place after data collection. Whereas Minimize forbids the collection of needless information, Hide suggests that any personal data that is processed should be hidden from plain view.

- *Separate:* The approach of the privacy strategy Separate is to process any personal information in a distributed fashion if possible. Thus, interrelationships between personal data vanish in contrast to a centralized processing.

- *Aggregate:* When implementing Aggregate, personal information is processed at a high level of aggregation. This level should only be so high as to remain useful, however. Details that are not needed for the functionality of the service vanish. This process could include statistical aggregation such that the details of identities are blurred.

- *Inform:* The privacy strategy Inform states that data subjects should be adequately informed whenever personal information is processed.

- *Control:* A common requirement of software systems is that data subjects should be in control of the processing of their personal information.

- *Enforce:* Enforce states that a privacy policy that is compatible with legal requirements should be in place and should be enforced.

- *Demonstrate:* The privacy strategy Demonstrate demands that data controllers are able to demonstrate compliance with their privacy policy and any applicable legal requirements. A good example fora pattern implementing this strategy is the use of audits.

Given the advances in different fields of ML ranging from image recognition, language translation, it becomes increasingly important for research scientists to be able to explore how the data is being interpreted by the models. Some of these advanced AI algorithms enable artificial agents to directly influence their environment through actions, such as moving a robot arm based on camera inputs or clicking a button in a web browser. While artificial agents have the potential to be increasingly helpful to people, current methods are held back by the need to receive detailed feedback in the form of frequently provided rewards to learn successful strategies.

In contrast, complex tasks require decision making at all levels. Within this regard, some algorithms train a manager policy to propose subgoals within the latent space of a learned world model and train a worker policy to achieve these goals. From predicted trajectories of model states, the algorithm optimizes two policies: The *manager* chooses a new goal every fixed number of steps, and the *worker* learns to achieve the goals through low-level actions.

All components are optimized concurrently, so the manager learns to select goals that are achievable by the worker. The manager learns to select goals to maximize both the task reward and an exploration bonus, leading the agent to explore and steer towards remote parts of the environment.

The data needed to train ML systems comes in a form that computers don't immediately understand. To translate concepts understand naturally by human-beings (e.g. words, sounds, or videos) to a form that the algorithms can process, embeddings are used. These refer to a mathematical vector representation that captures different facets (dimensions) of the data.

Clicking on any point brings up a list of nearest points and distances, which shows which words the algorithm has learned to be semantically related. This type of interaction represents an important way in

which one can explore how an algorithm is performing.

As privacy design strategies can be used to categorize privacy patterns by their fundamental approach, the same holds for privacy dark strategies. Based on Hoepman's privacy strategies, the following privacy dark strategies have been identified:

- *Maximize:* The goal of this dark strategy is to collect an inappropriate amount of data so that the amount of personal data that is collected, stored, or processed is significantly higher than what is actually needed for the task.

- *Publish:* The dark strategy 'Publish' can be characterized by the requirement that personal data (not intended to be public) is not hidden from plain view. There is no mechanism in place to hide personal data from unauthorized access, such as encryption or access control.

- *Centralize:* Centralize is the dark strategy which enforces that personal data is collected, stored, or processed at a central entity. For instance, some cookies can be stored centrally by the flash plug-in on the file system and may not be restricted to a specific web browser.

- *Preserve:* This dark strategy requires that interrelationships between different data items should not be affected by processing. They should rather be preserved in their original state for analysis instead of storing them in a processed form, e.g., aggregation.

- *Obscure:* In this dark strategy it is almost impossible for data subjects to learn how their personal data is collected, stored, and processed due to the use of a privacy policy with many technical terms, which might be difficult to understand for the average user.

- *Deny:* Patterns making use of this dark strategy make a data subject lose control of their personal data. With this dark strategy, a service provider can prevent users from taking actions that oppose that service provider's interest.

- *Violate:* This occurs if a privacy policy presented to the user is intentionally violated. A privacy policy is in place, yet it is intentionally not kept. As the users are unaware of the violation; the trust put into that service is not impacted at all.

- *Fake:* This dark strategy 'Fake' means that an entity collecting, storing, or processing personal data claims to implement strong privacy protection but in fact only pretends to.

Feng et al developed a design space of five key dimensions along with a "taxonomy to categorize, evaluate, and communicate different privacy choice design options. According to the authors, there are five key dimensions for the design space in privacy choices: type, functionality, channel, timing, and modality. Each of these dimensions have multiple options to be chosen from, and the controller should evaluate what options can help mitigate data subjects' informational vulnerabilities. Below is an image from Feng et *al*.'s research illustrating the multiple aspects of each of the five key dimensions in the design space for privacy:

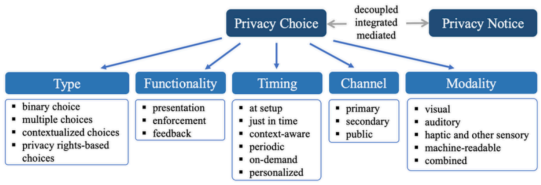

Figure 1.0 Feng et al.'s Design Space for Privacy

Building upon this framework of design space, the next section will describe the conceptual framework in more detail.

III. CONCEPTUAL FRAMEWORK

On the table below, common practices and the model-based changes are summarized as discussed above, including the type of informational vulnerability involved (Table 1.):

Type of Informational Vulnerability	Current Practices	TbD-Based Practices
Risk of wrong audience choice	Allow audience choice before every post, but do not make it mandatory.	Require data subjects to manually select the audience of the post before every post; after the data subject has selected the desired audience, show on average how many people will see the post and ask whether the data subject wants to change the desired audience.
Risk of unwanted spread of posted content	Do not warn data subjects that any posted content can potentially be re-uploaded indefinitely and on different platforms.	Warn data subjects before posting that other people could download or screenshot the content and repost it on Facebook or other platforms, therefore it might be impossible to retrieve control over its availability.
Risk of unwanted attention to posted content	Do not warn data subjects about the consequences of posting "publicly." Do not warn data subjects that a Facebook profile, as well as any content publicly posted is searchable on search engines such as Google (unless selected otherwise in the privacy settings). Do not warn	Warn data subjects before posting that by posting publicly, anyone on the internet can interact with the content. Actively ask data subjects if they want their profile to be indexed by search engines. Ask the data subject if he or she wants to allow comments or
	data subjects that even posting to friends only might generate unwanted attention from distant or estranged acquaintances.	shareability for the post. how information is shared with connected apps and to personalize advertising.
Risk of unwanted use (or unwanted consequences) of posted content or personal data	Do not inform data subjects that the content posted will be used for personalized advertising and also shared with third parties' apps.	Inform data subjects before posting that the content posted will be used by advertisers and third-party apps.
Risk of data protection harm due to lack of adequate support	Absence of support on data protection issues.	Provide an easily accessible 24/7 channel (i.e., an intelligent chat bot) to offer tech support to data subjects on data protection issues. In case the chat bot cannot provide a satisfactory answer, provide a feedback form that should be answered in due time.

Table 1.0 Strategies for informational vulnerabilities

Based on Table 1.0, the main guideline to be followed by UX designers is that there should be no black box, no prior data protection knowledge should be expected from data subjects, all data collection points should be transparent, clear and the explanation should be contextual and visceral, as close to the collection point as possible.

Model Development

During the model design, user sequences can be compared against an expert designer's sequence to facilitate interpretation of clusters and identification of those users who need aid. This design process can also leverage interactive visualization to make it easier for the designer to connect the sequence patterns to the clusters and an adjustable algorithm that takes input from the human analyst.

The stakeholder may not possess the technical experience necessary to interact with the algorithm, in which case working in tandem with a data scientist is recommended. However, this method is tool agnostic, meaning that it can be implemented with any visualization that meets the usability requirements, including one with an interface that provides graphical user interface (GUI) inputs for algorithm manipulation.

Step 1: Data Processing

The digital environment must be instrumented to record users' actions at a granular action-to-action level as they work through a problem or task. Any logged data not included in the mapping can be filtered out. If the data is still too complex, that is, there are too many actions, it is recommended that stakeholders develop an abstraction for the data, converting the lower-level actions to higher-level actions or behaviors. This process will help make the data easier for a human-being to read and understand.

Finally, once filtering and abstraction have been implemented, user data is converted into sequences of actions. Action sequences can be created based on timing, with actions recorded at set intervals, or they can be based on ordering.

All of these steps can be accomplished through data-processing scripts. An overview of the process has been shown in Figure 2.0.

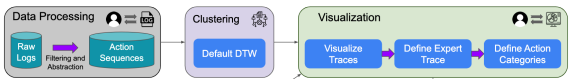

Figure 2.0 System-level view of user pattern identification and analysis

Step 2: Clustering

Once the data has been processed and converted into sequences of actions, the next step is to use a default algorithm to cluster these sequences by setting up a script that takes in the set of sequences as an input and runs the algorithm. In a dynamic programming manner, the algorithm fills in a matrix with values based on these distances as it steps through both sequences.

Step 3: Visualization

Visualization permits a holistic view of the entire population, making it easier to identify community-level patterns. By syncing a visualization of clusters produced by an algorithm with a visualization of user sequences, it is possible to learn how the algorithm understands the data.

Using the visualization, the stakeholder can categorize the actions that appear in the user sequences. For example, the action of re-reading a piece of text may belong to “clarification activities,” where a user seeks to reaffirm their understanding of a topic, while editing an answer may belong to “adjustment activities,” where a user adjusts their responses based on new information.

This process assumes that the stakeholder is familiar enough with the digital environment to either know or reasonably assume that engagement with a given action indicates a given content strategy (clarification, adjustment, etc.).

Step 4: Identification of Dark Patterns

Once the categories are set, the stakeholder can leverage their expert knowledge to analyze and compare sequences against each other and the expert trace. In doing so, they can identify the general characteristics for each set of clustered user sequences. By means of a relevant algorithm, the stakeholder identifies a set of sequences that are clustered near each other. Each sequence is then analyzed to identify high-level characteristics of how the user moved between action categories.

Step 5: Algorithmic Update and Iteration

At this stage, the stakeholder can update the algorithm based on their knowledge of the patterns in each cluster. In a process similar to what is described by Javvaji and colleagues (2020), the stakeholder can update the value added by the algorithm when there is a mismatch. These values are referred to as weights. Specifically, while the default algorithm adds a weight of one to any mismatch, the stakeholder can use their understanding of how users are interacting with the action categories to penalize actions within certain categories by defining a specific, greater value as the weight that should be added if a mismatch includes the designated action.

If sequence length is a confounding factor in the clustering process, weights can be adjusted such that, in situations where one sequence has ended but the other continues, a reduced weight can be applied for each step that the longer sequence continues.

CONCLUSION

This study explores the use of a design model to overcome dark patterns in which user sequences can be compared against an expert designer’s sequence to facilitate interpretation of clusters and identification of those users who need aid.

REFERENCES

[1] G. Eason, B. Noble, and I.N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] Abdi, S., Khosravi, H., & Sadiq, S. (2020). Modelling learners in crowdsourcing educational systems. In International Conference on Artificial Intelligence in Education (pp. 3–9). Springer.

[3] Abdi, S., Khosravi, H., Sadiq, S., & Darvishi, A. (2021). Open learner models for multi-activity educational systems. Artificial Intelligence in Education, 11–17. https://doi.org/10.1007/978-3-030-78270-2_2

[4] Ahmad, N., & Bull, S. (2008). Do users trust their open learner models? In International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (pp. 255–258). Springer.

[5] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., & Chatila, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.

[6] Ashenafi, M. M. (2017). Peer-assessment in higher education-twenty-first century practices, challenges and the way forward. Assessment & Evaluation in Higher Education, 42, 226–251.

[7] Carless, D., & Boud, D. (2018). The development of user feedback literacy: Enabling uptake of feedback. Assessment & Evaluation in Higher Education, 43, 1315–1325.

[8] Cho, K., & MacArthur, C. (2011). Learning by reviewing. Journal of Educational Psychology, 103, 73–84.

[9] Darvishi, A., Khosravi, H., & Sadiq, S. (2020). Utilising learner sourcing to inform design loop adaptivity. In European Conference on Technology Enhanced Learning (pp. 332–346). Springer.

[10] Darvishi, A., Khosravi, H., & Sadiq, S. (2021). Employing peer review to evaluate the quality of user generated content at scale: A trust propagation approach. In Proceedings of the Eighth ACM Conference on Learning@ Scale (pp. 139–150). Association for Computing Machinery

[11] Gardner, J., Brooks, C., & Baker, R. (2019). Evaluating the fairness of predictive user models through slicing analysis. In Proceedings of the 9th international conference on learning analytics & knowledge (pp. 225–234). Association for Computing Machinery.

[12] Gašević, D., Kovanović, V., & Joksimović, S. (2017). Piecing the learning analytics puzzle: A consolidated model of a field of research and practice. Learning: Research and Practice, 3, 63–78.

[13] Gyamfi, G., Hanna, B. E., & Khosravi, H. (2021). The effects of rubrics on evaluative judgement: A randomised controlled experiment. Assessment & Evaluation in Higher Education, 47(1), 126–143. https://doi. org/10.1080/02602938.2021.1887081

[14] Hadwin, A., Järvelä, S., & Miller, M. (2017). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In Handbook of self-regulation of learning and performance (pp. 83–106). Routledge.

[15] Han, Y., Wu, W., Yan, Y., & Zhang, L. (2020). Human-machine hybrid peer grading in SPOCs. IEEE Access, 8, 220922–220934.

[16] Hassan, T. (2019). Trust and trustworthiness in social recommender systems. In Companion Proceedings of The 2019 World Wide Web Conference (pp. 529–532). Association for Computing Machinery.

[17] Henderson, M., Phillips, M., Ryan, T., Boud, D., Dawson, P., Molloy, E., & Mahoney, P. (2019). Conditions that enable effective feedback. Higher Education Research & Development, 38, 1401–1416.

[18] Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and anovas. Frontiers in Psychology, 4, 863.

[19] Lee, W., Huang, C. H., Chang, C. W., Wu, M. K. D., Chuang, K. T., Yang, P. A. and Hsieh, C. C. (2018) Effective quality assurance for data labels through crowdsourcing and domain expert collaboration. In 21st International Conference on Extending Database Technology, EDBT 2018 (pp. 646–649). OpenProceedings.org.

[20] Levy, H., & Robinson, M. (2006). Stochastic dominance: Investment decision making under uncertainty (Vol. 34). Springer

[21] Matcha, W., Gašević, D., & Pardo, A. (2019). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. IEEE Transactions on Learning Technologies, 13(2), 226–245.

[22] Moon, T. (1996). The expectation-maximization algorithm. IEEE Signal Processing Magazine, 13, 47–60.

[23] Napoles, C., Sakaguchi, K., Post, M., & Tetreault, J. (2015). Ground truth for grammatical error correction metrics. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers) (pp. 588–593). Association for Computational Linguistics (ACL).

[24] Negi, S., Asooja, K., Mehrotra, S., & Buitelaar, P. (2016). A study of suggestions in opinionated texts and their automatic detection. In Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics (pp. 170–178). Association for Computational Linguistics

[25] Purchase, H., & Hamer, J. (2018). Peer-review in practice: Eight years of Aropä. Assessment & Evaluation in Higher Education, 43, 1146–1165.

[26] Ramachandran, L., Gehringer, E. F., & Yadav, R. K. (2017). Automated assessment of the quality of peer reviews using natural language processing techniques. International Journal of Artificial Intelligence in Education, 27, 534–581.

[27] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 3982–3992). Association for Computational Linguistics

[28] Topping, K. J. (2010). Peers as a source of formative assessment. In Handbook of formative assessment (pp. 73–86). Routledge.

[29] Urena, R., Kou, G., Dong, Y., Chiclana, F., & Herrera-Viedma, E. (2019). A review on trust propagation and opinion dynamics in social networks and group decision making frameworks. Information Sciences, 478, 461–475.

[30] Wang, W., An, B. and Jiang, Y. (2018) Optimal spot-checking for improving evaluation accuracy of peer grading systems. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 32, No. 1). AAAI Press.

[31] Wang, W., An, B., & Jiang, Y. (2020). Optimal spot-checking for improving the evaluation quality of crowdsourcing: Application to peer grading systems. IEEE Transactions on Computational Social Systems, 7, 940–955.

[32] Wind, D. K., Jørgensen, R. M., & Hansen, S. L. (2018). Peer feedback with peergrade. In ICEL 2018 13th International Conference on e-Learning (p. 184). Academic Conferences and Publishing Limited.

[33] Wright, J. R., Thornton, C., & Leyton-Brown, K. (2015). Mechanical TA: Partially automated high-stakes peer grading. In Proceedings of the 46th ACM Technical Symposium on Computer Science Education (pp. 96–101). Association for Computing Machinery.

[34] Xiong, W., & Litman, D. (2011). Automatically predicting peer-review helpfulness. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (pp. 502–507). Association for Computational Linguistics.

[35] Yang, M., Tai, M., & Lim, C. P. (2016). The role of e-portfolios in supporting productive learning. British Journal of Educational Technology, 47, 1276–1286.

[36] Yang, T.-Y., Baker, R. S., Studer, C., Heffernan, N., & Lan, A. S. (2019). Active learning for user affect detection. In Proceedings of the 12th International Conference on Educational Data Mining, EDM 2019, Montréal, Canada, July 2-5, 2019. International Educational Data Mining Society

(IEDMS) 2019 (pp. 208–217). Université du Québec; Polytechnique Montréal.

[37] Yeager, D. S., Purdie-Vaughns, V., Garcia, J., Apfel, N., Brzustoski, P., Master, A., Hessert, W. T., Williams, M. E., & Cohen, G. L. (2014). Breaking the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide. Journal of Experimental Psychology: General, 143, 804–824.

[38] Yu, F.-Y., & Wu, C.-P. (2011). Different identity revelation modes in an online peer-assessment learning environment: Effects on perceptions toward assessors, classroom climate and learning activities. Computers & Education, 57, 2167–2177.

[39] Zheng, L., & Huang, R. (2016). The effects of sentiments and co-regulation on group performance in computer supported collaborative learning. The Internet and Higher Education, 28, 59–67.

[40] Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. Higher Education Research & Development, 37, 883–897.

[41] Zimmerman, B. J., Bonner, S., & Kovach, R. (1996). Developing self-regulated learners: Beyond achievement to self-efficacy. American Psychological Association.

[42] Zong, Z., Schunn, C. D., & Wang, Y. (2021). What aspects of online peer feedback robustly predict growth in users' task performance? Computers in Human Behavior, 124, 106924.