AKADÉMIAI KIADÓ

Journal of Behavioral Addictions

DOI:
10.1556/2006.2024.00032


COMMENTARY

# User-avatar bond as diagnostic indicator for gaming disorder: A word on the side of caution

## Commentary on: Deep learning(s) in gaming disorder through the user-avatar bond: A longitudinal study using machine learning (Stavropoulos et al., 2023)

ALEXANDRE INFANTI[1†**], ALESSANDRO GIARDINA[2†], JOSIP RAZUM[3], DANIEL L. KING[4], STEPHANIE BAGGIO[5,6], JEFFREY G. SNODGRASS[7], MATTHEW VOWELS[2,13], ADRIANO SCHIMMENTI[8], ORSOLYA KIRÁLY[9,10], HANS-JUERGEN RUMPF[11], CLAUS VÖGELE[1] and JOËL BILLIEUX[2,12*]

[1] Department of Behavioural and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[2] Institute of Psychology, University of Lausanne, Lausanne, Switzerland

[3] Ivo Pilar Institute of Social Sciences, Zagreb, Croatia

[4] College of Education, Psychology & Social Work, Flinders University, Australia

[5] Institute of Primary Health Care (BIHAM), University of Bern, Bern, Switzerland

[6] Population Health Laboratory (#PopHealthLab), University of Fribourg, Fribourg, Switzerland

[7] Department of Anthropology and Geography, Colorado State University, Fort Collins, CO 80523-1787, USA

[8] Department of Human and Social Sciences, UKE – Kore University of Enna, Cittadella Universitaria, Enna, Italy

[9] Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary

[10] Centre of Excellence in Responsible Gaming, University of Gibraltar, Gibraltar, Gibraltar

[11] University of Lübeck, Department of Psychiatry and Psychotherapy, Lübeck, Germany

[12] Centre for Excessive Gambling, Addiction Medicine, Lausanne University Hospitals (CHUV), Lausanne, Switzerland

[13] The Sense Innovation and Research Center, CHUV, Lausanne, Switzerland

Received: February 28, 2024 • Revised manuscript received: April 18, 2024 • Accepted: April 28, 2024

†Alexandre Infanti and Alessandro Giardina equally contributed to this paper and are willing to share first authorship.

*Corresponding author.
E-mail: joel.billieux@unil.ch

**Corresponding author.
E-mail: infanti.alexandre@gmail.com

ABSTRACT

In their study, Stavropoulos et al. (2023) capitalized on supervised machine learning and a longitudinal design and reported that the User-Avatar Bond could be accurately employed to detect Gaming Disorder (GD) risk in a community sample of gamers. The authors suggested that the User-Avatar Bond is a “digital phenotype” that could be used as a diagnostic indicator for GD risk. In this commentary, our objectives are twofold: (1) to underscore the conceptual challenges of employing User-Avatar Bond for conceptualizing and diagnosing GD risk, and (2) to expound upon what we perceive as a misguided application of supervised machine learning techniques by the authors from a methodological standpoint.

KEYWORDS

machine learning, gaming disorder, user-avatar bond, classification, diagnosis

AKJournals

We commend Stavropoulos et al. (2023) for their study which aimed to test whether Gaming Disorder (GD) risk cases could be accurately detected based on Machine Learning (ML) algorithms trained with, among other variables, information regarding the User-Avatar Bond (UAB) (Blinka, 2008). Using longitudinal data, they claimed that the UAB has the potential to detect GD risk with implications for treatment and assessment. Specifically, the authors concluded that capitalizing on their method would permit the use of the UAB as a potential diagnostic indicator of GD risk. This kind of study is particularly relevant at this time, given the limited number of longitudinal studies, and the need to refine and improve the assessment and screening of GD. However, given the novelty of this approach and its potential impact on the field, we believe that some of the claims made by the authors warrant caution, both at the theoretical and methodological level.

In line with the authors' proposal, we agree on the psychological relevance of the relationship with the avatar in the study of problematic gaming patterns (Lemenager, Neissner, Sabo, Mann, & Kiefer, 2020; Razum & Huić, 2023). Observing such a relationship seems to be especially important in the presence of identity vulnerabilities such as poor self-esteem and self-concept clarity when Massively Multiplayer Online Role-Playing Games (MMORPGs) are played (Green, Delfabbro, & King, 2021; Király, Koncz, Griffiths, & Demetrovics, 2023; Szolin, Kuss, Nuyens, & Griffiths, 2022). Certainly, in clinical contexts involving individuals exhibiting problematic gaming behaviors, the examination of avatar perception could be a valuable avenue for gaining insight into implicit identity processes that underlie prevalent themes, conflicts, and developmental issues during consultations (Lemenager et al., 2020). Nevertheless, we believe that the authors' claim that *"the UAB could operate as a diagnostic indicator of GD risk both at present and prospectively (six months later), when addressed using trained ML/AI procedures"* (Stavropoulos et al., 2023, p.13) is premature.

Therefore, in this commentary, our objectives are twofold: (1) to underscore the conceptual challenges of employing UAB for conceptualizing and diagnosing GD risk, and (2) to expound upon what we perceive as a misguided application of supervised ML techniques by the authors from a methodological standpoint.

## CONCEPTUAL CRITICISM

The first reason for exercising caution is conceptual in nature. Although fascinating, Stavropoulos et al.'s (2023) idea that avatars might be considered as "digital phenotypes" (i.e., a digital/gamified footprint of an individual's mental health) is challenging for several reasons. First, digital phenotyping should provide data that is superior to self-report, and can use digital markers (Montag & Rumpf, 2021). In this sense, the objective analysis of in-game activities may provide many clues about the risk of addictive behavior (Larrieu, Fombouchet, Billieux, & Decamps, 2023), while current measures and conceptualizations of the UAB lack sufficient discriminatory power to be considered objective digital markers. Second, the concept of avatars as "digital phenotypes" requires that a relationship with the avatar exists. The existence of such a relationship may depend on two intertwined factors: 1) the type of videogame played and 2) the way avatars are experienced by the player. As for the first, in most MMORPGs the establishment of a meaningful relationship with an avatar is indeed possible and commonly documented, yet not intrinsically central (Mancini, Imperato, & Sibilla, 2019, 2024). However, for other types of games equally associated with GD and more popular nowadays, such as First Person Shooters (FPS), Real Time Strategy (RTS) games, Battle Royale (BR) games or Multiplayer Online Battles Arenas (MOBA), avatars are not central to game play and experience and can be customized only to a limited extent (Statista, 2023). Such constraints may diminish the likelihood of identification with or idealization of avatars, thus limiting players in fostering meaningful connections with these virtual representations and reducing their usefulness in understanding problematic gaming patterns (Király et al., 2023; Rehbein, King, Staudt, Hayer, & Rumpf, 2021). In Table 1, we propose an approximate inter-genre classification of most popular online games' genres based on the salience of avatars for the category,[1] i.e., the possible degree of avatars' customization in the category and the relevance for the gameplay/player experience.

A few differences are summarized here. In MOBA games, players are required to select from a predetermined roster of "heroes," resulting in limited or absent avatar customization compared to MMORPGs. Nonetheless, the choice of a hero in MOBA games, each characterized by distinct attributes and backgrounds, significantly influences gameplay dynamics. Moreover, MOBA players often develop emotional attachments to specific heroes, sometimes prioritizing their selection over strategic considerations for

*Table 1.* Game-play experience of avatars based on game genres

| | Customization | Relevance | Score |
|---|---|---|---|
| MMORPGs | High | Medium | 5 |
| MOBA | Low | High | 4 |
| BR | Medium | Low | 3 |
| FPS | Medium | Low | 3 |
| RTS | Very low | Very low | 0 |
| Sport games | Very low | Very low | 0 |

*Note*: MMORPGs = Massively Multiplayer Role-Playing Games; MOBA = Multiplayer Online Battle Arenas; BR = Battle Royale; FPS = First Person Shooter; RTS = Real-Time Strategy; Very Low = 0; Low = 1; Medium = 2; High = 3; Relevance = Impact on the gameplay/emotional bond of players; Score = summarized score of the values in the Customization and Relevance columns.

[1]We are aware that this classification represents a simplification of the huge variety of videogames (and avatars' settings) within the same genre. However, we believe it stays sufficiently true to the general features of each genre and to the inter-genre comparison.

individual matches. In FPS and BR games, there exists a degree of customization, such as altering weapon appearances or selecting the avatar from predefined "skins." However, the customizations in these genres tend not to confer competitive advantages in gameplay, and the bond between players and avatars tends to be more aesthetic-instrumental rather than emotionally driven. Lastly, in RTS and sports games (with certain exceptions depending on the sport), individual avatars are absent, with players instead choosing from groups represented as teams or factions.

It is evident that the genre of the video game may impose certain important constraints on the avatar-player relationship. However, the mere classification of video game genres does not ensure a specific perception of the avatar. Embedded within the preference for a particular game genre is thus the players' individual experiences with avatars, which can vary in nature. Stavropoulos et al. (2023) base their proposal on the players' experience of avatars as *extensions of themselves* into the virtual world – thereby suggesting processes of identification with the avatar, idealization of the avatar, and/or utilization of the avatar within the game environment to compensate for personal and interpersonal deficiencies. Nevertheless, problematic gaming can also occur when avatars are experienced as mere tools to interact with the game or as friends and adventures' companions (Snodgrass et al., n.d; Green et al., 2021). For example, according to Banks (2015) the level of psychological differentiation of players from their avatars (i.e., the autonomy of avatars from players themselves) is only one of four factors determining the UAB. The others include the level of emotional investment, the ability to imagine avatars as something more than just digital tools or personalized entities (i.e., a suspension of disbelief) and the degree of perceived control over the avatar. Based on how these elements vary, Banks and Bowman (2016, 2021) propose that players can relate to the avatar: (a) as an *object*, where avatars are experienced in a non-social way, i.e., as mere tools to play the game; (b) as *me*, where a significant emotional bond sustains the identification with a non-idealized avatar; (c) as a *symbiote*, where there is an identification with an idealized avatar; or (d) as *other,* where avatars are perceived as separate being in a social and emotionally salient way, and thus are akin to friends or adventures' companions. According to this conceptual framework, the approach advocated by Stavropoulos et al. (2023) may effectively identify problematic gaming behaviors in instances where players exhibit strong emotional connections with their avatars, as seen in "me" or symbiote avatars. This approach, however, may fall short in detecting problematic gaming when there is a lack of emotional attachment between the players and their avatars, as observed in the "avatar as an object" category, or when the avatar is perceived as a socially significant entity distinct from the player, as exemplified in the avatar "as other" category (Snodgrass et al., n.d.). These considerations might also help explain Stavropoulos et al.'s (2023) finding that the immersion dimension of the UAB Questionnaire (UAB-Q; Blinka, 2008) was the best predictor of GD risk in their sample. One reason behind such a stronger association may be that the items of the current UAB-Q immersion dimension (1) mostly refer to thinking about the character or the game while not playing (recalling the "preoccupation" criterion of the DSM-5 Internet Gaming Disorder condition; American Psychiatric Association, 2013; Castro-Calvo et al., 2021), but also that (2) they assess a general emotional bond with the character (i.e., "sometimes I feel ashamed for/proud of my character"). Accordingly, these items do not necessarily refer to the experience of the avatar as an extension of the self (i.e., as "me" or as a symbiote) but they might also imply a perception of the avatar as a sort of "playmate" (i.e., the condition of avatar as "other"). In this respect, it is noteworthy that Blinka, Sirinkova & Stasek (2023) recently tested an updated version of the UAB-Q, the UAB 2.0, on 6,391 adult gamers. In this revised version, the dimension which showed the highest correlation ($\beta = 0.32$) with GD symptoms was the compensation of gamers' weaknesses through the avatars' superior characteristics. Furthermore, in the UAB 2.0 an Emotional Bond dimension was identified via factor analysis, which could be an important variable for examining other kinds of UABs in which avatars are perceived as "*other*" (Banks, 2015; Banks & Bowman, 2021).

In summary, scenarios exist in which avatars are perceived as extensions of the self fostering identification, idealization, or compensation and contributing to GD symptoms, as it is sometimes observed in MMORPGs or MOBA players (Stavropoulos et al., 2023; Szolin et al., 2022). However, there could be also multiple scenarios in which problematic gamers have an avatar that is perceived as a separate companion (as in the "*other*" category proposed by Banks (2015), or even cases where no particular emotional bond is created between the player and the avatar, as it could happen with FPS or RTS games (Rehbein et al., 2021; Snodgrass et al., n.d.). From this perspective, the UAB as implemented by Stavropoulos et al. (2023) may hold clinical significance in instances where avatars are perceived as "me" or as a symbiote (e.g., within MMORPGs). Nevertheless, it seems premature to consider the UAB as an inherently reliable indicator for GD diagnosis universally. An indicator must provide a clear threshold which would be a fundamental step to be taken to go in this direction. Furthermore, additional research is warranted to investigate whether specific UABs correspond with various video game genres (Banks & Bowman, 2021).

## METHODOLOGICAL CRITICISM

The second point of caution we emphasize is of a methodological nature. The way ML algorithms are implemented in Stavropoulos et al. (2023), but also more recently in Brown et al. (2024) and Hein, Conkey-Morrison, Burleigh, Poulus, and Stavropoulos (2024), is based on an elevated proportion of simulated (i.e., algorithm-generated) data. A crucial step in ML pertains to the splitting of the available database into two different sets: the *train set,* which is used to fit the model, and the *test set,* which is used to evaluate the fitted model on unseen data to estimate its performance (Rosenbusch, Soldner, Evans, & Zeelenberg, 2021). To obtain an equal proportion of No-GD risk and Yes-GD risk cases in their two sets, Stavropoulos et al. (2023) generated virtual data (i.e., simulated gamers profiles) using an algorithm called K-NN Synthetic Minority Oversampling Technique (SMOTE).[2] This approach is particularly useful since it tackles a common problem in psychological research, where the clinical group usually represents a minority of the population, leading to a considerable imbalance in databases. Under such circumstances, a specific ML classifier model would use the majority class (non-clinical population) for its predictions and give very limited importance to the minority class (clinical population; Chawla, Japkowicz, & Kotcz, 2004). Thus, by using the K-NN SMOTE algorithm, Stavropoulos et al. (2023) adopted a potentially sound approach to bypass this issue. Nevertheless, these authors implemented the algorithm *before* splitting their data to produce the train and test sets. By using K-NN SMOTE, the authors artificially inflate the number of cases in the minority group (Yes-GD risk). Surprisingly, the authors also artificially inflated, instead of under-sampling, the number of cases in the majority group (No-GD risk). After the use of the K-NN SMOTE algorithm, the final database used was composed of 1,060 participants, where 424 Yes-GD risk cases (80% of this subsample) and 100 No-GD risk cases (18.87% of this subsample) were algorithm-generated data.[3] Crucially, these algorithm-generated data represent 49.43% of the final sample before the split is made to create the *train* and *test* sets. It is worth noting that the SMOTE algorithm has been criticized for its inability to generate reliable cases in the minority class (Kosolwattana et al., 2023). Moreover, a related problem is that algorithm-generated data are present in the *test* set used to establish the accuracy of the fitted model (i.e., the *test* set is composed of a mixture of real and simulated data). These decisions are questionable because the specific way in which the authors have augmented their dataset with synthetic data negatively impacts the generalizability of the model and significantly inflates its apparent performance. In our view, it would have been important that Stavropoulos et al. (2023) fully disclose that the methodology they implemented might be able to detect "mainly algorithm-generated data" and that further research is needed to establish the actual validity of this method as a potentially valid diagnostic indicator in the context of real cases.

We argue that a sounder approach could be to implement the K-NN SMOTE algorithm *after* splitting the data and exclusively in the *train set*. This would render the *test set* realistic and implies that the model's accuracy is tested in a real condition (see Fig. 1 for a graphical explanation).

As the database used by Stavropoulos et al. (2023) is not available in the online supplement, we used an available dataset to illustrate our proposal (Table 2). The database we used for this purpose is available from the open science framework: https://doi.org/10.17605/OSF.IO/2P6SX. Our proposal was thus operationalized using a large dataset in which participants with or without a mental health condition completed a self-reported scale measuring various impulsivity traits (the short French UPPS-P impulsive behavior scale, see Billieux et al., 2012 for the scale and Billieux et al., 2021 for more details on the sample). The database comprises 18,953 participants, and among them, 385 have a mental disorder (clinical cases). We compared the approach of Stavropoulos et al. (2023) and the alternative proposal in the present comment (i.e., implementing the K-NN SMOTE algorithm after the splitting of the data and on the *train* set exclusively) to predict the clinical status of the participants based on the UPPS-P questionnaire assessing impulsivity traits.[4] For the supervised ML analyses, we used the Random Forest ensemble model, which was the most accurate in the study by Stavropoulos et al. (2023), but without tuning. We aimed to demonstrate the potential impact of including algorithm-generated data inside the *test* set. Thus, our comparison is focused on this very point, which is methodological and not specific to a dataset.

The procedure is illustrated in Fig. 1. Table 2 compares the accuracy of the two approaches using real cases from an available database. Impulsivity traits (negative urgency, positive urgency, lack of premeditation, lack of perseverance, and sensation seeking) were used as predictors of the clinical status (non-clinical or clinical). The (diagnostic) accuracy, which represents the percentage of correct prediction, was 99% with the approach used by Stavropoulos et al. (2023) and 98% with the method we suggest in the present paper. The accuracy itself, however, is not sufficient to assess the quality of the model's predictions. For that reason, metrics such as precision (the model's ability to prevent false positive predictions) and recall (or sensitivity, the model's ability to identify positive results accurately) are also reported for a more nuanced evaluation of the

[2]This algorithm generates simulated data for the minority class (oversampling technique) while taking into account a number (K) of nearest neighbors (NN) when considering the Euclidean distance. This algorithm can also randomly remove/select some cases from the majority class (undersampling) to balance the data. For more details about the algorithm please see Chawla, Bowyer, Hall, and Kegelmeyer (2002).

[3]The authors have acknowledged the oversampling of the Yes-GD risk cases. However, they did not mention the oversampling of the No-GD risk cases.

[4]The supplementary material provided by Stavropoulos et al. (2023) does not include the bake recipe, folds train boot or VIP, which is susceptible to errors and compromises the reproducibility of the procedure. Also, several unclear manipulations (e.g., creating folds for the cross-validation without using them, or the use of another SMOTE algorithm in the recipe) have been found in the provided code. For this reason, and to guarantee the reproducibility of the present analyses and findings, we adapted the code and the procedure to illustrate our proposal. Our user-friendly data analytic code is available in supplementary material. The issues we encountered further highlights the importance of endorsing open science practice where well-documented and reproducible analytical code are available (Eben et al., 2023).

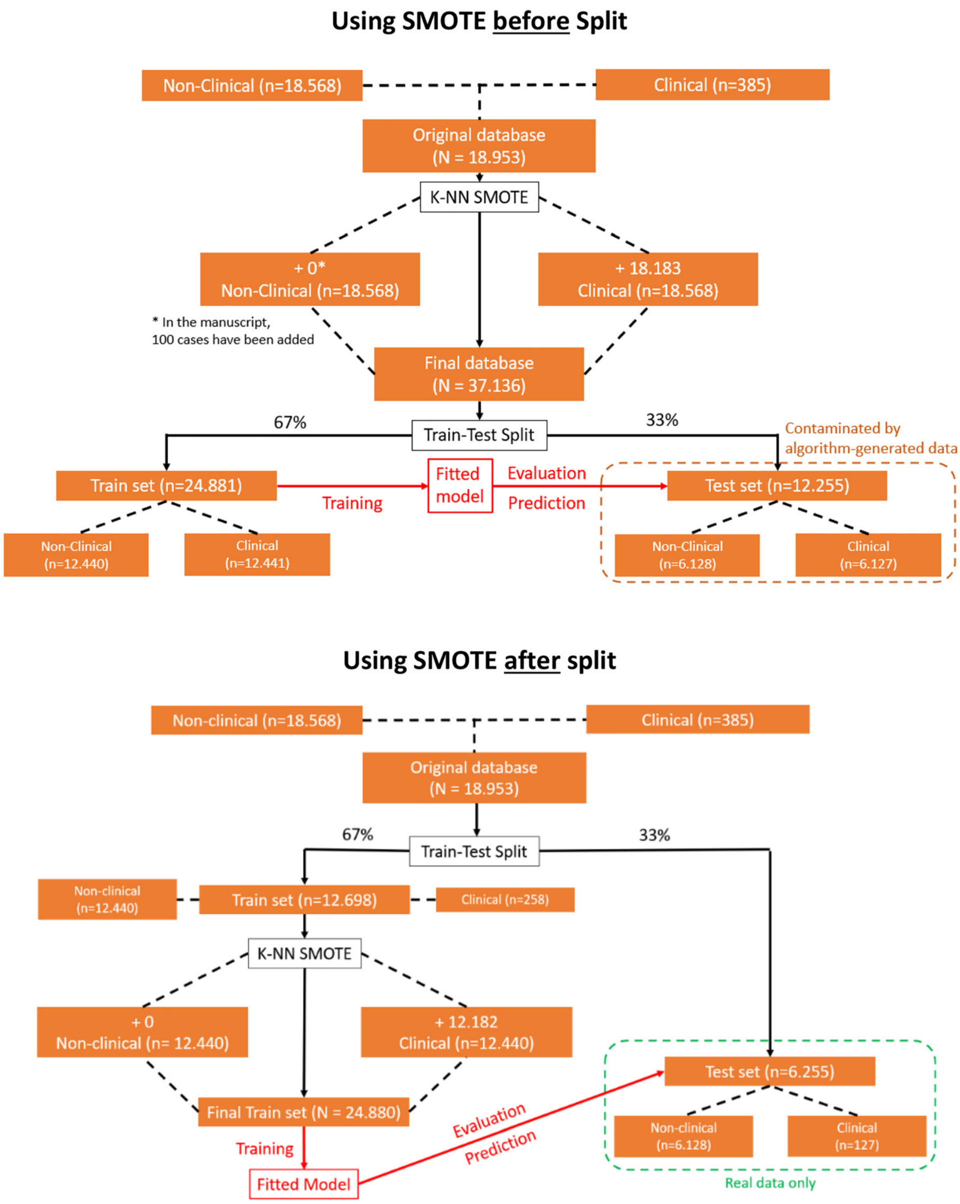


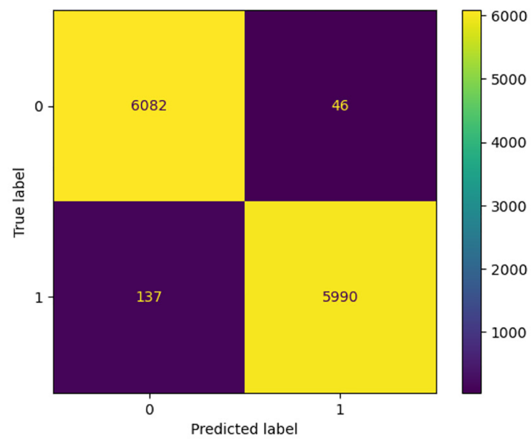
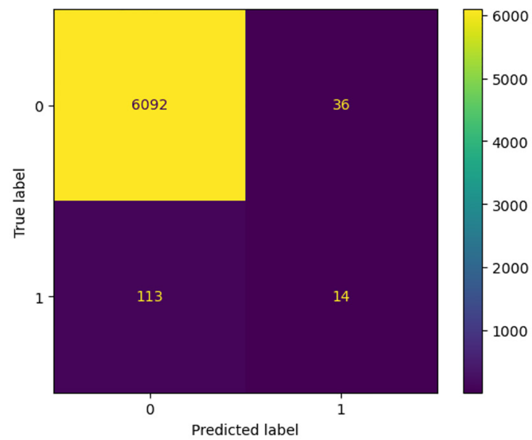
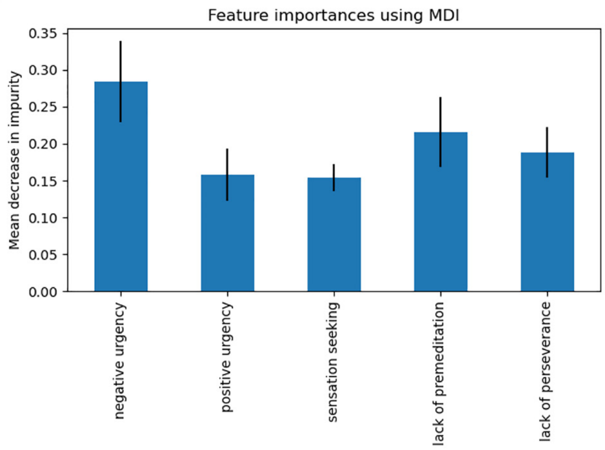
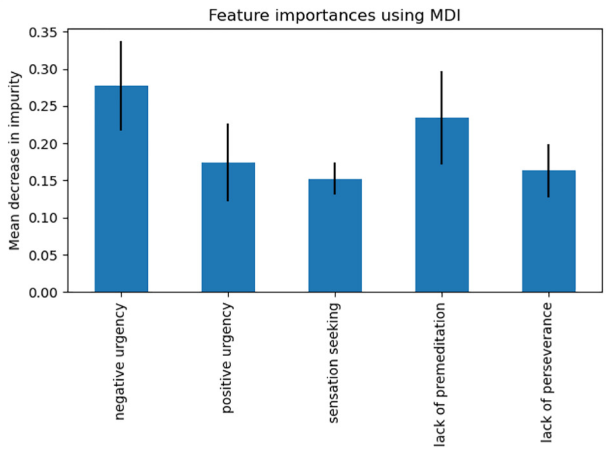

*Fig. 1.* Difference between the two methods

models. In our comparison, we noticed that when simulated data was generated after the sample split, precision and recall scores dropped significantly for clinical sample predictions. Precision decreased from 99% to 28%, and recall decreased from 98% to 11%, leading to a very poor predictive model. Our result thus challenges the practical relevance and utility of the model proposed by Stavropoulos et al. (2023).

It would also have been beneficial for the authors to consider implementing a supervised ML regression analysis. The GDT-4 scale was primarily designed to assess the severity of disordered gaming by using a total score rather than providing a diagnosis (Pontes et al., 2021). This is even more relevant when considering the impact of the data quality on a supervised ML model's performance. The prediction of a supervised ML model is, in the best-case scenario, as accurate as the instrument output (Fardouly, Crosby, & Sukunesan, 2022). Regarding this point, it is worth noting that, in Stavropoulos et al. (2023), the functional impairment criterion was not considered necessary to identify participants as Yes-GD risk cases.[5] This approach contrasts with the recommendation provided by Pontes et al. (2021), which consists of meeting all criteria (a criterion being endorsed when answering "Often" or "Very often") to identify disordered gamers, referring to the conservative

[5]The analytic code included in the online supplement does not match with what the authors say they have done. Indeed, while the authors say in their article that a response modality of 4 ("often") or higher is the rule to consider a gaming disorder criterion as endorsed, their data analytic code considers a criterion endorsed when the response modality is 3 ("sometimes") or higher.

*Table 2.* Comparison of the impact of data's split before or after the generation of data

<table>
<tr><th>Generating data before train-test split</th><th>Generating data after train-test split</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
<table>
<tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr>
<tr><td>0</td><td>0.98</td><td>0.99</td><td>0.99</td><td>6128</td></tr>
<tr><td>1</td><td>0.99</td><td>0.98</td><td>0.98</td><td>6127</td></tr>
<tr><td>accuracy</td><td></td><td></td><td>0.99</td><td>12255</td></tr>
</table>
</td><td>
<table>
<tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr>
<tr><td>0</td><td>0.98</td><td>0.99</td><td>0.99</td><td>6128</td></tr>
<tr><td>1</td><td>0.28</td><td>0.11</td><td>0.16</td><td>127</td></tr>
<tr><td>accuracy</td><td></td><td></td><td>0.98</td><td>6255</td></tr>
</table>
</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Precision = the proportion of detection of true values among the predicted values, computed using the formula (true positive)/(true positive + false positive); Recall = The capacity of the model to find the true value, computed using the formula (true positive)/(true positive + false negative).

approach to diagnosis defended in the ICD-11 (Billieux et al., 2017). Thus, the nature of the sample identified as GD risk gamers remains unclear, leading to potentially highly involved but healthy gamers being included in this sample (Billieux, Flayelle, Rumpf, & Stein, 2019). For this reason, we believe that it would have been helpful if Stavropoulos et al. (2023) had strictly followed Pontes et al.'s (2021) recommendations to strengthen diagnostic-related claims, especially when creating groups based on the GDT-4 (Pontes et al., 2021).

## CONCLUSION

The UAB may be an important element to explore and consider in the context of case formulation for people with problematic gaming behaviors. Nevertheless, we believe that there are often limits to the clinical relevance of the UAB. We are not convinced at this time that the UAB concept is appropriately supported by empirical evidence to be considered a clinical feature or diagnostic indicator of GD. Moreover, the results brought by the authors through their methodology do not provide sufficiently robust arguments to support the UAB for that purpose. We are also concerned about the generalization of the results based on how supervised ML was implemented in their study (see Brown et al., 2024 and Hein et al., 2024 for other recent studies using a similar methodology).

In conclusion, based on the current state of literature, the relevance of the UAB in GD can vary significantly depending on the interaction between the game genre and the way avatars are experienced by the player. Determining the relevance of UAB in any given case is unlikely to be a straightforward process. Furthermore, the results obtained by Stavropoulos et al. (2023) are limited to the identification of algorithm-generated – and thus simulated – data for the Yes-GD risk case, which hinders the generalization of the results to actual problematic gamers. When generating data after the sample split, we observed a significant decrease in the model's ability to detect clinical cases. Therefore, the model performance was greatly impacted when testing the model on a sample consisting solely of actual cases. Further case studies, research on clinical samples focusing on the relationship with the avatars in different game genres, and the evaluation of different methods to assess the relationship with avatars are, therefore, needed before exploring further the idea of the UAB as a "digital phenotype" or a potential indicator (or diagnostic feature) of GD.

*Funding sources:* This work is part of the DRIVEN project funded by the Luxembourg National Research Fund under the PRIDE program (PRIDE17/ 12252781).

*Authors' contribution:* Alexandre Infanti and Alessandro Giardina wrote the first draft of the paper, under the supervision of Joël Billieux. Alessandro Giardina took the lead on the conceptual criticism (first part of the comment), while Alexandre Infanti took the lead on methodological and statistical criticism (second part of the comment). Alexandre Infanti computed and analyzed the supervised machine learning analyses. Josip Razum, Daniel L. King, Jeffrey G. Snodgrass, Adriano Schimmenti, Orsolya Király, Hans-Juergen Rumpf, and Claus Vögele provided input regarding the conceptual criticism. Stéphanie Baggio and Matthew Vowels provided input regarding the methodological and statistical criticism. All authors contributed to the writing and editing of the manuscript.

*Conflict of interest:* Joël Billieux, Daniel King, and Hans-Jürgen Rumpf are Associate Editors for the Journal of Behavioral Addictions. Orsolya Király is editorial board member for the Journal of Behavioral Addictions. The authors report no other potential conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## SUPPLEMENTARY MATERIAL

Supplementary data to this article can be found online at https://doi.org/10.1556/2006.2024.00032.

## REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Banks, J. (2015). Object, me, symbiote, other: A social typology of player-avatar relationships. *First Monday*. https://doi.org/10.5210/fm.v20i2.5433.

Banks, J., & Bowman, N. D. (2016). Emotion, anthropomorphism, realism, control: Validation of a merged metric for player–avatar interaction (PAX). *Computers in Human Behavior*, *54*, 215–223. https://doi.org/10.1016/j.chb.2015.07.030.

Banks, J., & Bowman, N. D. (2021). Some assembly required: Player mental models of videogame avatars. *Frontiers in Psychology*, *12*, 701965. https://doi.org/10.3389/fpsyg.2021.701965.

Billieux, J., Flayelle, M., Rumpf, H.-J., & Stein, D. J. (2019). High involvement versus pathological involvement in video games: A crucial distinction for ensuring the validity and utility of gaming disorder. *Current Addiction Reports*, *6*(3), 323–330. https://doi.org/10.1007/s40429-019-00259-x.

Billieux, J., Heeren, A., Rochat, L., Maurage, P., Bayard, S., Bet, R., … Baggio, S. (2021). Positive and negative urgency as a single coherent construct: Evidence from a large-scale network analysis in clinical and non-clinical samples. *Journal of Personality*, *89*(6), 1252–1262. https://doi.org/10.1111/jopy.12655.

Billieux, J., King, D. L., Higuchi, S., Achab, S., Bowden-Jones, H., Hao, W., … Poznyak, V. (2017). Functional impairment matters in the screening and diagnosis of gaming disorder: Commentary on: Scholars' open debate paper on the World Health Organization ICD-11 Gaming Disorder proposal (Aarseth et al.). *Journal of Behavioral Addictions*, *6*(3), 285–289. https://doi.org/10.1556/2006.6.2017.036.

Billieux, J., Rochat, L., Ceschi, G., Carré, A., Offerlin-Meyer, I., Defeldre, A. C., … Van der Linden, M. (2012). Validation of a short French version of the UPPS-P impulsive behavior scale. *Comprehensive psychiatry*, *53*(5), 609–615. https://doi.org/10.1016/j.comppsych.2011.09.001.

Blinka, L. (2008). The relationship of players to their avatars in MMORPGs: Differences between adolescents, emerging adults and adults. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *2*(1), Article 5. https://cyberpsychology.eu/article/view/4211/3252.

Blinka, S. & Stasek. (2023, August). *User-avatar bond 2.0: Development of a new-old scale and its association with gaming disorder and life satisfaction*. 8th international conference on behavioral addictions (ICBA 2023), Incheon, South Korea. *Journal of Behavioral Addictions*, *12*(Supplement-1), 1–365. https://doi.org/10.1556/2006.2023.00800.

Brown, T., Burleigh, T. L., Schivinski, B., Bennett, S., Gorman-Alesi, A., Blinka, L., & Stavropoulos, V. (2024). Translating the user-avatar bond into depression risk: A preliminary machine learning study. *Journal of Psychiatric Research*, *170*, 328–339. https://doi.org/10.1016/j.jpsychires.2023.12.038.

Castro-Calvo, J., King, D. L., Stein, D. J., Brand, M., Carmi, L., Chamberlain, S. R., … Billieux, J. (2021). Expert appraisal of criteria for assessing gaming disorder: An international Delphi study. *Addiction*, add.15411. https://doi.org/10.1111/add.15411.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953.

Chawla, N. V., Japkowicz, N., & Kotcz, A. (2004). Editorial: Special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, *6*(1), 1–6. https://doi.org/10.1145/1007730.1007733.

Eben, C., Bőthe, B., Brevers, D., Clark, L., Grubbs, J. B., Heirene, R., … Billieux, J. (2023). The landscape of open science in behavioral addiction research: Current practices and future directions. *Journal of Behavioral Addictions*, *12*(4), 862–870. https://doi.org/10.1556/2006.2023.00052.

Fardouly, J., Crosby, R. D., & Sukunesan, S. (2022). Potential benefits and limitations of machine learning in the field of eating disorders: Current research and future directions. *Journal of Eating Disorders*, *10*(1), 66. https://doi.org/10.1186/s40337-022-00581-2.

Green, R., Delfabbro, P. H., & King, D. L. (2021). Avatar identification and problematic gaming: The role of self-concept clarity. *Addictive Behaviors*, *113*, 106694. https://doi.org/10.1016/j.addbeh.2020.106694.

Hein, K., Conkey-Morrison, C., Burleigh, T. L., Poulus, D., & Stavropoulos, V. (2024). Examining how gamers connect with their avatars to assess their anxiety: A novel artificial intelligence approach. *Acta Psychologica*. https://doi.org/10.1016/j.actpsy.2024.104298.

Király, O., Koncz, P., Griffiths, M. D., & Demetrovics, Z. (2023). Gaming disorder: A summary of its characteristics and aetiology. *Comprehensive Psychiatry*, *122*, 152376. https://doi.org/10.1016/j.comppsych.2023.152376.

Kosolwattana, T., Liu, C., Hu, R., Han, S., Chen, H., & Lin, Y. (2023). A self-inspected adaptive SMOTE algorithm (SASMOTE) for highly imbalanced data classification in healthcare. *BioData Mining*, *16*(1), 15. https://doi.org/10.1186/s13040-023-00330-4.

Larrieu, M., Fombouchet, Y., Billieux, J., & Decamps, G. (2023). How gaming motives affect the reciprocal relationships between video game use and quality of life: A prospective study using total playtime indicators. *Computers in Human Behavior*, *147*, 107824. https://doi.org/10.1016/j.chb.2023.107824.

Lemenager, T., Neissner, M., Sabo, T., Mann, K., & Kiefer, F. (2020). "Who Am I" and "how should I Be": A systematic review on self-concept and avatar identification in gaming disorder. *Current Addiction Reports*, *7*(2), 166–193. https://doi.org/10.1007/s40429-020-00307-x.

Mancini, T., Imperato, C., & Sibilla, F. (2019). Does avatar's character and emotional bond expose to gaming addiction? Two studies on virtual self-discrepancy, avatar identification and gaming addiction in massively multiplayer online role-playing game players. *Computers in Human Behavior*, *92*, 297–305. https://doi.org/10.1016/j.chb.2018.11.007.

Mancini, T., Imperato, C., Sibilla, F., & Musetti, A. (2024). Can personal identity protect against problematic gaming? A study on the relationships between identity motives, user–avatar bond, and problematic gaming in a sample of MMORPG players. *Identity*, *24*(1), 64–78. https://doi.org/10.1080/15283488.2023.2291644.

Montag, C., & Rumpf, H.-J. (2021). The potential of digital phenotyping and mobile sensing for psycho-diagnostics of Internet use disorders. *Current Addiction Reports*, *8*(3), 422–430. https://doi.org/10.1007/s40429-021-00376-6.

Pontes, H. M., Schivinski, B., Sindermann, C., Li, M., Becker, B., Zhou, M., & Montag, C. (2021). Measurement and conceptualization of gaming disorder according to the world health organization framework: The development of the gaming disorder test. *International Journal of Mental Health and Addiction*, *19*(2), 508–528. https://doi.org/10.1007/s11469-019-00088-z.

Razum, J., & Huić, A. (2023). Understanding highly engaged adolescent gamers: Integration of gaming into daily life and motivation to play video games. *Behaviour & Information Technology*, 1–23. https://doi.org/10.1080/0144929X.2023.2254856.

Rehbein, F., King, D. L., Staudt, A., Hayer, T., & Rumpf, H.-J. (2021). Contribution of game genre and structural game characteristics to the risk of problem gaming and gaming disorder: A systematic review. *Current Addiction Reports*, *8*(2), 263–281. https://doi.org/10.1007/s40429-021-00367-7.

Rosenbusch, H., Soldner, F., Evans, A. M., & Zeelenberg, M. (2021). Supervised machine learning methods in psychology: A practical introduction with annotated R code. *Social and Personality Psychology Compass*, *15*(2), e12579. https://doi.org/10.1111/spc3.12579.

Snodgrass, J. G., Sagstetter, S., Branstrator, J. R., Giardina, A., Zhao, K. X., Lacy, M. G., … Billieux, J. (n.d). Player-avatar relationships and gaming experience: Personal bonds with characters do not indicate higher risk of disordered play, (working manuscript).

Statista (2023). *Most popular video game genres among internet users worldwide as of 3rd quarter 2022, by age group*. statista. https://www.statista.com/statistics/1263585/top-video-game-genres-worldwide-by-age/#:~:text=Shooter%20games%20ranked%20as%20the,across%20almost%20all%20age%20groups.

Stavropoulos, V., Zarate, D., Prokofieva, M., Van De Berg, N., Karimi, L., Gorman Alesi, A., … Griffiths, M. D. (2023). Deep learning(s) in gaming disorder through the user-avatar bond: A longitudinal study using machine learning. *Journal of Behavioral Addictions*, *12*(4), 878–894. https://doi.org/10.1556/2006.2023.00062.

Szolin, K., Kuss, D., Nuyens, F., & Griffiths, M. (2022). Gaming disorder: A systematic review exploring the user-avatar relationship in videogames. *Computers in Human Behavior*, *128*, 107124. https://doi.org/10.1016/j.chb.2021.107124.