

RESEARCH

Open Access



Online gambling forums as a potential target for harm reduction: an exploratory natural language processing analysis of a reddit.com forum

Mark van der Maas^{1*} and Jim Samuel¹

Abstract

Objectives Globally, there has been a rapid increase in the availability of online gambling. As online gambling has increased in popularity, there has been a corresponding increase in online communities that discuss gambling. The movement of gambling and communities interested in gambling to online spaces presents new challenges to harm reduction. The current study analyses a forum from a popular online forum hosting website (reddit.com) to determine its suitability as a source for data to inform gambling harm reduction in online spaces.

Methods The current study provides an exploratory analysis of 1,141 unique posts and 11,668 comments collected from the online forum r/onlinegambling. The dataset covers posts and comments from August 5, 2015, to October 30, 2023. Natural language processing (NLP) techniques were used to identify common terms and phrases, identify topics with high rates of participant engagement and perform a sentiment analysis of posts and comments.

Results Sentiment analysis results showed that the majority of posts and comments were positive, but there were substantial numbers of negative and neutral content. Positive content was often congratulatory and focused on winning, neutral posts more commonly focused on practical advice, and negative posts were more commonly concerned with avoiding operators perceived as illegitimate by forum participants.

Conclusions Results from this study show that there is a varied and robust discussion of different aspects of gambling on r/onlinegambling. Our exploratory analyses suggest that this reddit forum provides important information on how users communicate motivations to gamble, interpretations of gambling experiences, and define potential harms related to gambling online as well as how to avoid or remedy those harms. Reddit forums discussing gambling have great potential for future research interested in more specific aspects of harm reduction and prevention related to online gambling, particularly when using NLP techniques.

*Correspondence:

Mark van der Maas
mark.vandermaas@rutgers.edu

¹Rutgers, The State University of New Jersey, New Brunswick, USA



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Online gambling markets are rapidly expanding globally; between 2008 and 2022, revenue from online gambling grew from \$21.7B [1] to \$63.5B [2]. This trend is driven in part by improved 24-hour access to gambling for anyone with a computer or mobile phone and by the rapid legalization of gambling activities in jurisdictions of varying levels around the world. This expansion has been particularly rapid in the US, where the repeal of the Professional and Amateur Sports Protection Act (PASPA) in 2018 has led to the legalization of sports wagering in 38 of 50 states and the District of Columbia in just six years [3].

In response to this rapid growth, there has been growing interest in gambling harm reduction and prevention research. Broadly defined, harm reduction is an approach that attempts to reduce the harm from a behavior rather than focusing on reducing the behavior itself [4]. As noted in the review by Marrionneau et al. [5], most current strategies for gambling harm reduction were designed with land-based gambling in mind. Their call-to-action notes multiple challenges that online gambling environments present including broader availability, targeted marketing, and difficulties in regulation. However, they also note that the features of online space present opportunities for harm reduction. The expansion of online gambling opportunities has corresponded with the growth of online gambling communities, which foster socializing, knowledge gathering, and the sharing of perceptions, attitudes, and feelings relevant to their gambling participation [6]. Understanding these communities is important because they may constitute the initial or primary source of information on gambling activities for novice gamblers. While some reviews on the topic of online gambling communities note the value of online spaces discussing problem gambling [5, 7, 8], there is less focus on the potential for online gambling interest groups in harm reduction. Studies of drug use communities suggest that there is a wide range of risk management practices that are shared informally between people who use drugs that meaningfully contribute to the goals of harm reduction [9]. For example, people selling drugs have been found to be an important resource for people who use opioids but wish to reduce their risk of overdose by accidental use of fentanyl [10]. Their expertise and gate-keeper status to the wider risky drug market makes them an invaluable source of information for people who continue to use drugs but want to do so relatively safely.

A growing body of research suggests there is cause for concern in online communities organized around gambling. A recent systematic review [7] found that while these communities can offer support in the experience of negative gambling consequences, overall, they have the effect of motivating participation in gambling activities. One study of Finnish adolescents and young adults found

that daily online gambling community participation was associated with compulsive internet use and higher problem gambling scores [11]. A related study reported that a stronger sense of belonging to an online community was also associated with higher problem gambling scores [12]. There is some evidence that online communities are particularly attractive to individuals who experience issues with mental health [13], those who have comparatively little support in their offline lives [14], and those who experience persistent negative emotional states such as loneliness [11]. Social media are significant sources of misinformation relevant to various health concerns [15, 16]. Similarly, online communities using social media have the potential to spread misinformation or reproduce potentially harmful narratives related to gambling.

Nearly 70 years ago, Howard Becker's groundbreaking research on cannabis users showed how important informal networks were to the initiation of cannabis use, interpretation of the drug's effects, and how to avoid unwanted consequences of cannabis use [17]. Contemporary research has found communities of use continue to be vital to the safety of their members. Informal information networks are important to ensure that illicit drugs purchased come from safe sources and that risk is minimized when in the act of buying. For example, in a sample of methamphetamine users in the US state of Oregon, information regarding the potential presence and dangers of fentanyl in methamphetamines was shared through informal networks. This information led to changes in practices including increasing their scrutiny of who supplied their drugs and changes in how they used them [18]. There are similar findings among users of opioids [10, 19] and cannabis [20]. These communities of use can also serve as useful targets for harm reduction intervention. A systematic review of mental health interventions using social networking sites targeting people below the age of 25 found that such interventions had high engagement and usability, and were useful in improving mental health literacy [21].

Textual discussions and exchanges in online forums offer a wealth of data for analysis. However, the sheer volume of data can be challenging. Researchers are increasingly turning to NLP methods, which include an array of useful exploratory analyses, algorithmic approaches, tools and libraries for efficient insights discovery, and Large Language Models (LLMs), as effective approaches, to help deal with important aspects of that challenge [22 {Samuel, 2024 #847}]. NLP methods, including LLMs, enable researchers to process and analyze vast quantities of text more effectively, offering insights into trends, sentiment, and conversations without relying on manual effort [23 {Rahman, 2021 #849}]. NLP methods have challenges, for example - LLMs can be computationally inefficient. Furthermore, while these NLP methods and

models can streamline data analysis, their use — especially in high-stakes fields like behavioral health can pose significant dangers. For example there are issues in the reliability and accuracy of NPL approaches, transparency in the output generated by such models, and problems of accountability when such models cause harm [24]. A simulated study where physician communications to patients were compared with those drafted by LLMs and those assisted by LLMs found that the LLM draft responses improved subjective efficiency in 76.9% of cases. However, LLM responses also led to a risk of severe harm in 7.1% of cases and death in 1 case out of 156 [25]. In another example, in a test where they were asked “What to do if someone is not breathing?”, two popular chatbot programs powered by LLMs only met the Resuscitation Council United Kingdom guidelines 9.5% and 11.4% of the time [26]. Such issues can be partially mitigated through active supervision by a knowledgeable human user, but still must be taken seriously when applying LLMs when addressing mental health issues.

Discussion of information and education in existing frameworks of gambling harm reduction typically centers on larger campaigns aimed at the general population and product warning labels. There is limited evidence of the effectiveness of these measures for gambling harm prevention [27]. However, online environments offer an opportunity to provide tailored messaging to those for whom the information is most relevant [5]. Further, there is also evidence that social media users can be useful in mental health promotion [28]. This suggests that interventions engaging with moderators or influential contributors on reddit forums may be helpful in ensuring that information reducing risky practices is accurate and visible to users. For example, when Reddit quarantined communities like r/TheRedPill and r/The_Donald, it significantly reduced new member growth by 79.5% and 58%, respectively. This shows that moderation strategies can help limit the spread of harmful content. However, these measures did not reduce existing users’ levels of misogyny or racism, highlighting that while moderation can restrict community growth, it may not change entrenched behaviors [39]. Similarly, studies on online mental health communities found that moderators play a crucial role in fostering supportive spaces for individuals to share sensitive experiences and seek help. Effective moderation not only ensures the quality and credibility of information but also creates safe and empathetic environments (40, 41).

Given the growing size of these communities and the increasing focus on online gambling from both regulated and unregulated gambling providers, it is crucial to understand the nature and scope of gambling related discussions on online forums. The current study provides an exploratory analysis of a large online forum that discusses

online gambling. The purpose of these exploratory analyses was to discover prospective behavioral insights and to determine the potential value of Reddit forums as a target for future analyses.

Methods

Dataset description

Data for the current project were collected from the r/onlinegambling forum (or “subreddit”) on the popular message board hosting website reddit.com. The subreddit is a monitored forum that describes itself simply as an online gambling community. R/onlinegambling was established on September 6, 2008, and boasts roughly 24,000 subscribers. Along with hosting the discussions of community members, the site also provides recommendations for online gambling websites, links to other gambling focused subreddits, a list of rules for the community to follow, and resources for those experiencing gambling-related problems including the National Council on Problem Gambling Helpline (United States), Gamblers Anonymous, The Center for Addiction and Mental (Canada), Gamcare (United Kingdom) and Gambling Help Online (Australia).

Data were collected via Reddit’s API, specifically targeting the “online gambling” subreddit (<https://www.reddit.com/r/onlinegambling/>). The Python Reddit API Wrapper (PRAW) facilitated smooth interaction with Reddit’s platform. The data extraction process followed these steps: (1) Initializing the Reddit API Connection: To begin, a Reddit client was established using PRAW (Python Reddit API Wrapper). This required inputting essential credentials that allowed the script to access the Reddit API in read-only mode. (2) Fetching Posts: Data was extracted from top, hot, new, and rising posts in the “online gambling” subreddit. Key details such as post title, text, ID, score, number of comments, URL, and creation time were captured and stored in a structured format for analysis. (3) Fetching Comments: For each post, comments were retrieved, with truncated threads being excluded. Both the comment text and associated post IDs were collected, and pauses were introduced during the process to comply with API rate limits.

The post data was consolidated, eliminating duplicates, and corresponding comments were added to create a complete dataset that included both post metadata and comments. This data collection methodology was carefully implemented to adhere to API usage restrictions while effectively gathering substantial amounts of information for subsequent analysis.

The final scraped dataset includes a total of 1,141 unique posts and 11,668 comments. The dataset covers posts from as early as August 5, 2015, at 11:11:03, to the most recent on October 30, 2023, at 22:30:35, providing community engagement, and topical shifts within

the online gambling community over an eight-year span. Information on date and time, external links, reddit's internal "likes" system ("up-votes" and "down-votes"), and user provided flags ("flare").

Dataset preprocessing

The collected data were prepared for use with NLP tools and methods using numerous data cleaning processes. We identified spam patterns using iterative manual reviews of the data and by applying searches for common spam, bot and advertisement terms such as posts with links "www..." and high degree of unjustified repetition identified via algorithmic duplicate message identification and word frequency analysis methods. Similarly, we also identified additional stop words through manual iterations of unigrams and identifying words which did not contribute to the sensemaking process. We also used commonly known abusive words filters to search for and flag exclusively profane content. Such removal is algorithmic and non-exhaustive, because spelling mistakes and changes in spaces, punctuations or spellings escape common algorithmic searches. First, comments and posts that were suspected to be produced by non-human programs ("bots") and/or served as advertisements ("spam") were removed from the dataset. A total of 1,941 comments out of 11,668 comments were identified as spam and removed, and no posts were identified as spam. Second, posts were screened for offensive language. Such language included general profanity, racial and ethnic slurs, or sexual profanity. Most indications of offensive language were not removed from the dataset. A total fifty-two comments were removed where the post contained nothing but an offensive slur. Third, a collection of "stopwords" was applied to exclude common words to focus the descriptive analysis on gambling-related content. Commonly used terms in online conversations included words such as 'please', 'like', 'new', 'contact', 'questions', 'to', 'get', 'days', 'old', 'read', 'rules', 'remember', 'forget', 'would', 'thank', and variations of the word 'join'. We also identified Subreddit-specific terms like '/r/onlinegambling', 'moderators', 'message', 'subreddit', 'Discord', 'gg', 'dzcqv4p4dg', 'n', 't', 'r', 'onlinegambling', 'sidebar', 'subscribe', 'posting'. Terms related to automated posts or bot interactions included 'bot', 'action', 'performed', 'automatically', and 'compose'. The terms were identified as being indicative of automated posts by human team members experienced in using social media data. By removing these terms, the analysis was able to concentrate on more substantive and relevant words within the posts and comments. This step was crucial for accurately analyzing the most prominent topics and themes discussed by the community, eliminating noise, and focusing on terms that offer genuine insights into the behavior and preferences of users engaging with online gambling topics. In

the presentation of unigrams in the results section, the following phrases were removed: ', 'don', 'm', 've', 'll', 'u', 'Just', 'did', 'deleted'.

Analysis

Engagement analysis

To gauge the popularity of posts on the subreddit, an engagement score was calculated. The engagement is a simple composite score that includes the number of upvotes or down votes that a post received combined with the number of comments that a post received. Engagement analyses were summarized for topics of the most popular posts, and for different types of gambling discussed in posts. Human coders were provided a list of key words and exemplar posts under the collected topics and provided a summary term for each. Members of the research team involved in the identification of themes had expert knowledge in social media communications and gambling activities.

Sentiment analysis using VADER [on posts]

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. It uses a combination of a sentiment lexicon, which is a list of lexical features (e.g., words) which are labeled according to their semantic orientation as either positive or negative. When presenting unigrams classified by the sentiment analysis, words sharing a common root were grouped together (example: play, playing, played).

This study applied the following approach to sentiment scoring. Each piece of text (such as the text of the post) is analyzed to assess the sentiment expressed in it. The VADER tool evaluates the text to determine the amounts of positive, negative, and neutral sentiment it contains based on its lexicon of sentiment-laden words. The analysis yields scores for each category (positive, negative, neutral) and a composite score (compound) which aggregates these into a single measure. The compound score ranges from -1 (most negative) to +1 (most positive). Once each text has been assigned sentiment scores, the compound score is used to categorize the overall sentiment of the text as either positive, negative, or neutral [29]. A threshold is applied where: Scores equal to or greater than 0.05 are classified as 'Positive'. Scores equal to or less than -0.05 are classified as 'Negative'. Scores between -0.05 and 0.05 are classified as 'Neutral'.

Results

VADER sentiment analysis

The overall distribution of posts showed most posts indicated a positive sentiment (68.5%) with fewer negative (16.0%) and neutral (15.4%) posts. Table 1 displays the

Table 1 Twenty most common unigrams in posts as categorized by sentiment score

Positive	Count	Neutral	Count	Negative	Count
win*	568	casino*	27	site*	78
play*	530	bet*	25	casino*	69
casino*	477	gamble*	24	gamble*	55
game	341	com	23	account*	53
site*	287	game*	20	deposit*	52
money	283	got	12	play	52
gamble*	259	looking	12	money	50
slot*	244	sports	12	withdraw	47
withdraw*	226	try	12	win*	41
account	190	deposit*	11	bet*	36
say*	183	sites*	11	say*	33
paid*	178	stake	11	game*	32
know	177	day	10	want*	32
bets*	173	use	10	know	31
deposit*	171	slots	9	bank	30
player*	170	live	8	cash*	27
want	170	money	8	com	27
use*	159	vpn	8	pay*	26
card*	142	able	7	live	25
bonus*	141	paid*	7	verify*	25

* Denotes combined counts of phrases including a shared root word

Table 2 Twenty most common unigrams in comments as categorized by sentiment score

Positive	Count	Neutral	Count	Negative	Count
casino*	1649	casino*	209	casino*	508
play*	1577	use*	181	site*	306
win*	1360	deleted	157	play*	305
web*	1035	stake	143	use*	262
bonus	995	withdraw*	143	money	250
game*	944	site*	141	gamble*	231
money	741	game*	105	game*	211
deposit	656	gamble*	103	withdraw*	207
gamble*	650	com	98	lose*	200
withdraw*	629	try*	98	stake	197
bet*	624	crypto	83	shit*	186
slot*	623	cash*	80	win*	162
good	546	bet*	77	try*	153
stake	536	money	69	people	145
account*	501	vpn	67	deposit*	133
know	483	deposit*	63	know	127
player*	470	bovada	62	time	125
tried*	456	pay*	60	illegal	123
lol	445	account*	56	account	122
time	433	kyc	56	crypto	119
com	428	bitcoin	54	fuck	116
crypto	424	send*	52	vpn	111
paid*	415	slot*	52	cash*	105
cash*	404	know	49	kyc	102
make*	397	got	47	scam	102

* Denotes combined counts of phrases including a shared root word

most frequent twenty words used in each sentiment category. Regarding positive posts, many of the most common terms focus on core aspects of gambling, such as “casino”, “play”, “bet”, “game”, “win”, and “gambling”. Terms related to the financial aspects of online gambling were also popular including “withdraw”, “deposit”, “account”, and “money”. These groups of popular terms suggest a general focus on the activity itself and a discussion of management of the outcomes.

The most frequent terms in neutral sentiment posts showed some similar patterns as positive posts but with a greater focus on specific forms of gambling such as sports (indicating sports betting) and slots. The popularity of com also suggests that neutral posts are focused on more specific discussions such as particular online gambling websites or games. “Stake” is included but most commonly refers to a specific online gambling casino, Stake.com, which was the most discussed site on the forum.

Negative sentiment posts again show that popular gambling core terms, financially focused terms are the most common. Money, account, withdraw, bank, and deposit are all popular terms and indicate that negative sentiment posts more commonly focus on money. Both “bank” and “verify” are only found in the top twenty unigrams for negative posts.

Comments displaying a positive sentiment were most common (55.4%) followed by neutral (24.9%) and negative (19.7%) (Table 2). Positive comments showed similar words to those included in the positive sentiment analysis for posts and were more closely related to success. Unique comments in the top forty unigrams for Positive comments include “bonus”, “free”, “win” and “won”.

The most common words for neutral comments are similar to those found in positive comments. In comparison to both positive and negative comments, the top twenty neutral comments unigram shows unique phrases that indicate specific sites or products. “www” implies providing specific URLs to other users and words like “bovada” and “bitcoin” also refer to more specific products rather than the general phrases like “crypto” found in the positive and negative comments. Unique top forty unigrams for neutral comments also include “try”, “cash” and “withdraw” perhaps indicating more specific advice about interacting with sites. Neutral sentiment comments included words like crypto, VPN (virtual private network), and KYC (Know Your Customer). These terms are much more closely related to the specifics of online gambling and how to avoid restrictions and regulations related to online gambling. VPNs allow users to avoid regional restrictions for gambling sites, and Know Your Customer refers to the techniques that online gambling sites used to validate the identities of users.

The most common words used in the negative sentiment comments were similar to those found in the

positive comments. The most common words focus on themes of money and playing. There are several words that show in the top twenty-five most common words that are not present in the top words for positive comments. “illegal,” “scam,” “KYC” and “people” are notable differences between positive and negative comments. This suggests that many negative comments involve discussion of both illegitimate sites or practices and avoiding detection from sites that users are accessing against the sites’ regulations. The popularity of vulgarity as indicated by “shit” and “fuck” being in the top twenty-five unigrams also indicates abusive or inflammatory language in negative comments.

Some popular terms were discussed differently across sentiment categories. For example, “bonus*” was a common term used among discussions of wins. In these cases, bonuses led to some of the most dramatic wins in the posts. However, when bonuses were discussed in posts coded as neutral or negative, they were discussed with trepidation.

Negative: “They don’t need to rig the games. Deposit bonuses are literally designed so that there is a very high likelihood of you busting your deposit (and any winnings) before completing playthrough.”

Negative: “Punt Casino offers a no deposit “bonus”. I met the wagering requirements for the amount to become withdrawable. But they never sent my winnings. Contact customer service, turns out you can’t withdraw winnings from a no deposit “bonus” no matter what. It’s monopoly money, that’s it. Customer service admitted to it.”

Neutral: “Earn your cash points through real-money bets, not with bonus money or free spins, and exchange them for cash with no wagering requirements”.

Neutral: “I feel like most of these problems come when people use bonuses. Straight cash deposits are usually always paid out.”

For each of the above instances, users are discouraging players from using bonuses, portraying bonuses as ways that make tracking and accessing funds more difficult. In the case of the negative post, there is an implication that these are intentional tactics on the part of operators to encourage excessive betting. This was especially true in discussion of “playthrough” or “rollover” requirements to access bonuses. Of 686 comments where bonuses were discussed, 215 (31.3%) were cautionary.

Engagement analysis

To understand what people in the online gambling forum were most interested in, we first looked at the posts that got the most engagement. We noticed that the most popular posts often discussed certain common themes. These themes helped us create the main topics for the analysis.

After identifying these themes, we collected keywords mostly from the words used in the top ten posts and the top forty unigrams (the most common single words). These keywords were like the popular terms in the top engaged posts. There could be more synonyms too. Each post in the dataset was then analyzed to see if it contained one or more of these keywords in the title or content. If it did, the post was assigned to the corresponding topic. Posts that did not match any specific category were grouped under the miscellaneous topic. Our analyses focused on the topics that are most central to the experiences and interests of online gambling forum users.

The topic with the highest level of engagement was “losses” with an average engagement score of 45.00. Keywords included in this topic were “loss”, “lost”, and “down”. Comments responding to these posts typically focused on negative experiences with online casino sites. Users would often provide commiseration and advice on how to avoid issues in the future if it was perceived that a mistake or risky practice was indicated. Such advice focused on which online gambling providers to avoid and those that were considered more reliable.

Another trend in the discussion of losses was observed in a review of posts identified as containing abusive language by our analysis showed frequent blaming of the user for their negative experiences in online gambling. A total of thirty-two posts and 514 comments were flagged for abusive language. These instances were hand-sorted by a member of the team with expert knowledge in the field of gambling research to flag the attribution of blame for negative experiences with online gambling. There were eighty-four instances of blame being attributed to users for their negative experiences with online gambling among comments with abusive language. These comments typically focused on not understanding the terms of service of websites, trying to avoid regional restrictions, and misunderstanding odds or complaining about predictable losses. Examples are included below:

Dude if you deposit at an online casino that accepts payment in a fiat currency you deserve to get fucked, which by the way you are.

Degenerate gamblers: “insert any casino” is kinda fuckin scammy. Lost my whole paycheck with not one \$1000 win. Fuhcked up.

Yeah this seems like one of those prickly moves alright. Hate that, but if they have the T&Cs (Terms and Conidtions) they can fairly fuck you over as they wish.

Each of these comments are responses to different posts discussing issues with losses or withholding pay by an online casino. Along with attribution of blame, the use of stigmatizing language to discuss problem or excessive

gambling was also present. “degenerate” and “degen” were the most common pejorative used to describe excessive behavior, used in 12 instances in the posts, and 34 instances in the comments and “addict” and “addiction” were also used in 4 instances in the posts and 14 instances in the comments. In contrast, “problem gambling” which is a common less stigmatizing term for the experience of gambling harm was used 5 times and “Gambling Disorder”, the current preferred term to describe clinical levels of gambling harm, was used once.

The second highest engagement topic was casinos with an average engagement score of 37.92. Keywords included “casino”, “stake”, and “rollbit”. These posts focused on weighing the pros and cons of different online casino platforms and recommendations from community members for what they considered to be more legitimate casinos. For example, the keywords “stake” and “rollbit” are the names of popular online casinos at the time of publication. The third highest engagement topic was “wins” with a score of 33.49. Keywords included “win”, “cashed”, “max”, “won”, “jackpot”, and “profit”. The community members typically engaged in discussion of success with congratulations. For example, the post “Very happy with this win!” Received a post engagement score of 81. This post, celebrating a win, resonates with the community’s enthusiasm for sharing and congratulating success stories. The post showed an image taken from an online slots machine win of approximately \$1200 on a \$2.50 bet. Sharing images of big wins was a common occurrence for this theme. The fourth highest engagement topic was “advice” with a score of 29.92. Keywords for these posts included “advice”, “tips”, “help”, and “legit”. Community members were willing to engage with posts that asked specific questions or invited a topic of discussion around the best choices when online gambling. This advice included how to avoid online gambling providers that were considered to be unreliable, unwilling to pay wins, or otherwise deceive users or fix the outcome of games. General advice also included focusing on playing for the enjoyment of gambling rather than profit, budgeting responsibly, and how to avoid getting accounts suspended or terminated.

Engagement scores were also calculated for mentions of specific types of gambling. The most popular form of gambling was casino gambling with an engagement score of 34.50. Online casinos were the central form of gambling discussed on the forum. Slot machines showed the second highest engagement type of gambling with an average score of 26.53. While an online casino might entail many different individual games, the term slot machine tends to define a narrower set of game characteristics. Blackjack had the third highest engagement average engagement score with 25.08. Fourth was sports

betting and fifth was lottery with average engagement scores of 19.92 and 15.33, respectively.

Discussion

The purpose of our research was to provide an exploratory analysis of the content of the r/onlinegambling subreddit, discover prospective behaviors and to consider its suitability as a data source to inform harm reduction interventions. Relevant to the goals of harm reduction, the r/onlinegambling, our analyses showed that a core feature of the forum was informational in nature. The forum’s focus on advice and sharing of gambling experiences reflects existing research on online gambling communities [30]. Users were interested in sharing their experiences and looking for ways to gamble online that minimize their exposure to risk, largely from sites that were viewed as scams or how to avoid some of the practices perceived as predatory or exploitative on the part of the online casino operators. Discussions around reducing risk included discouragement of using gambling inducements offered by gambling site operators which have been shown to contribute to gambling intensity, frequency, and at-risk behaviors during online gambling [31]. Some of these features of the practices reflect the practices of people who use drugs and reflect some of the earliest research on drug use communities, namely Howard S. Becker’s *Marihuana Use and Social Control* [17]. This seminal study found that users of cannabis relied on informal networks to learn the motivations to use the drug, the techniques of use, and how to minimize risk while doing so. Reddit forums discussing gambling may present an opportunity for increasing literacy on both risky play habits and understanding Gambling Disorder as a mental health issue.

Due to the popularity of social media, there is a growing body of information on the value of these platforms. Our analyses show that users of the r/onlinegambling forum are both highly interested in sharing information and in discussing the accuracy or usefulness of that information. Web-based peer-supported forums for gambling or substance use disorders further demonstrate the potential of online spaces to provide mutual support, enabling individuals to overcome barriers like physical location or stigma while benefiting from the anonymity of digital platforms [42]. However, such interventions would need to be vetted carefully. In a study of the practices of moderators on the forum r/The_Donald, moderator interventions did reduce the activity of problematic contributors, but also resulted in a rise in toxicity and sharing of inaccurate or misleading information [32].

As noted in the call-to-action by Marionneau et al. [5], artificial intelligence approaches, particularly NLP, present an opportunity for studying harm reduction for online gambling. They note that these approaches to

data might be particularly useful for targeting messaging for online participants or flagging risky behaviors or users for potential intervention. Our analyses show that the data collected from reddit forums were appropriate for machine-learning techniques and would be suitable for more direct and theoretical analysis in the future. For example, our analyses showed that users were direct in asking for advice and tips and that responding reviewers were direct in providing it. It would be feasible and practical to identify users who are seeking information and what topics they are interested in learning about.

The exploratory analyses also showed potential areas for concern regarding the exchanges on the r/online gambling subreddit. One feature of the forum that shows a potential for increasing harm is the forum's strong focus on wins. Discussion of wins was the second most popular topic identified by our analysis of the text data. These posts would often have screen shots of large totals, a discussion of how quickly the total was amassed, and congratulatory responses from commenters. An overrepresentation of wins on the forum and the positive attention that they receive from users can have the effect of misleading users about the chances of winning large sums in online gambling. Similar mechanisms are found in gambling advertising [33] the presentation of winning large sums dominates messages. The focus on positive depictions of gambling is also likely to seriously limit the effect of messaging designed to help players limit their play [34].

Another potential challenge that these forums present for harm reduction is their role in perpetuating stigmatizing language or beliefs regarding gambling harm. Exchanges between users showed regular attribution of blame on users for negative experiences when gambling online. This trend was apparent in the review of vulgarities and bigoted language in posts and comments. There were many instances of abusive language where users showed a lack of understanding of the risks of the sites that they were gambling on, such as unregulated offshore online casinos, or poor and overly risky strategies or behaviors, such as claiming sites are rigged following a loss. In these instances, there were depictions of the users as being blameworthy for their own losses or difficulties with gambling. There also instances of highly stigmatizing pejorative terms to describe users who were described as gambling excessively or recklessly such as "degenerate" and "addict". Though these terms were relatively rare when considering the forums as a whole, they were far more common than terms to use to describe gambling that are considered less stigmatizing such as "Gambling Disorder" or "problem gambling". Such reactions to instances where users are expressing difficulties related to their gambling can be highly stigmatizing [35]. Gambling Disorder remains a highly stigmatized

condition and feelings of self-stigma can discourage those experiencing problems with gambling from seeking help [36]. While online environments can be useful in reducing the impact of stigma for treatment seekers [37], the current results suggest that social media platforms such as reddit can be important media for spreading stigmatizing language and attitudes [38, 39].

Limitations

We are aware of the limitations of this study, which are summarized below. However, while these limitations may hamper alternative goals, our objectives were not significantly impacted by these limitations and we intend to address and resolve some of the limitations in our future research. First, there are some constraints placed by reddit.com's API that limited the amount of retrievable data in single requests. Our data collection strategy included multiple collection requests and may have excluded some data between such requests. Relatedly, we were no able to include the content of posts or comments that were deleted or removed at the time of collection. This has the potential to bias the current sample, particularly where deleted posts used offensive language, which was more common among posts categorized as negative by our sentiment analyses. Our current study was also limited to studying a single subreddit related to gambling and as a result is limited to that particular subreddit (r/online-gambling). Subreddits focused on other games or aspects of gambling may yield very different results.

Our approach to using engagement and word frequencies to identify trends and gauge topics is well-grounded but limited by the fact that alternate topics may probably be missed when using this approach. Future research can mitigate this by using a multiple-methods strategy with established topic modeling libraries and LLMs.

There were also several analytical challenges posed by the above research plan. First, while we developed a protocol for identifying posts generated by a computer program ("bots"), it is possible that some of these posts were included in the analyses. There are also limits to the VADER program. The program is considered suitable for exploratory analyses but may struggle with complex language features like sarcasm or slang, leading to sentiment misclassification. Third, the user community of reddit.com has developed its own lexicon that may confuse analysis models, potentially resulting in content misinterpretation. Future analyses are planned to address some of the above issues in more topic-focused research.

Conclusions

The r/onlinegambling forum on reddit.com showed an active community of online gamblers who were primarily interested in exchanging practical information about their gambling behaviors. This information included a

discussion of how to avoid some of the harms and dangers of online gambling as defined by the users and how to deal with negative experiences or scenarios when encountered. Understanding how harm is defined and discussed by a community of users online engaged in potentially harmful behaviors is crucial to the development of effective harm reduction or prevention messaging. Reddit and similar online platforms represent great value for harm reduction related to gambling, both in understanding perceptions of harm in online spaces and as a potential site for future interventions.

Author contributions

MV wrote the main body of the introduction, discussion and conclusion and contributed to writing the methods section. MV also assisted in the interpretation of the analysis results. JS wrote the methods section, collected study data, created the Natural Language Processing Model and performed analyses. JS also supported on preparing the final manuscript. All authors reviewed the manuscript and provided edits for the final draft.

Funding

This research was supported by the Office of the Vice Provost for Research, Rutgers University.

Data availability

Data for the current study were collected from the public website reddit.com. Upon publication the study data will be made available through github.

Declarations

Competing interests

The authors declare no competing interests.

Conflict of interest

The authors declare no conflicts of interest in the publication of this study.

Consent to participate

Not applicable.

Human ethics

This study was approved by the Rutgers Electronic Institutional Review Board, project # 2021000040.

Received: 13 December 2024 / Accepted: 15 April 2025

Published online: 13 May 2025

References

1. Spectrum Gaming Group. Internet Gambling Developments in International Jurisdictions: Insights for Indian Nations 2010 [Available from: <https://web.archive.org/web/20120324092325/http://www.indiangaming.org/info/alerts/Spectrum-Internet-Paper.pdf>
2. Grand View Research. Online gambling market size, share & trends analysis report by type (Sports betting, casinos, poker, Bingo). By device (Desktop, Mobile), By Region (North America, Europe, APAC, Latin America, MEA). And Segment Forecasts; 2022. pp. 2023–30.
3. American Gaming Association. Interactive Map: Sports Wagering in the U.S. 2024 [Available from: <https://www.americangaming.org/research/state-gaming-map/>
4. Lenton S, Single E. The definition of harm reduction. *Drug Alcohol Rev*. 1998;17(2):213–9.
5. Marionneau V, Vuohio H, Karlsson N. Gambling harm prevention and harm reduction in online environments: a call for action. *Harm Reduct J*. 2023;20(1):92.
6. Boyd DM, Ellison NB. Social network sites: definition, history, and scholarship. *J computer-mediated Communication*. 2007;13(1):210–30.
7. Savolainen I, Sirola A, Vuorinen I, Mantere E, Oksanen A. Online communities and gambling behaviors—A systematic review. *Curr Addict Rep*. 2022;9(4):400–9.
8. Sirola A, Savela N, Savolainen I, Kaakinen M, Oksanen A. The role of virtual communities in gambling and gaming behaviors: A systematic review. *J Gambl Stud*. 2021:1–23.
9. Duff C. The importance of culture and context: rethinking risk and risk management in young drug using populations. *Health Risk Soc*. 2003;5(3):285–99.
10. Carroll JJ, Rich JD, Green TC. The protective effect of trusted dealers against opioid overdose in the US. *Int J Drug Policy*. 2020;78:102695.
11. Sirola A, Kaakinen M, Savolainen I, Oksanen A. Loneliness and online gambling-community participation of young social media users. *Comput Hum Behav*. 2019;95:136–45.
12. Savolainen I, Kaakinen M, Sirola A, Koivula A, Hagfors H, Zych I, et al. Online relationships and social media interaction in youth problem gambling: A four-country study. *Int J Environ Res Public Health*. 2020;17(21):8133.
13. Barak A, Boniel-Nissim M, Suler J. Fostering empowerment in online support groups. *Comput Hum Behav*. 2008;24(5):1867–83.
14. Cispke E, Horne O. Pro-eating disorder websites: users' opinions. *Eur Eat Disorders Review: Prof J Eat Disorders Association*. 2007;15(3):196–206.
15. Chou W-Y, Gaysynsky A, Cappella JN. Where we go from here: health misinformation on social media. *American Public Health Association*; 2020.
16. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. *J Med Internet Res*. 2021;23(1):e17187.
17. Becker HS. Marijuana use and social control. *Soc Probl*. 1955;3(1):9.
18. LaForge K, Stack E, Shin S, Pope J, Larsen JE, Leichtling G, et al. Knowledge, attitudes, and behaviors related to the fentanyl-adulterated drug supply among people who use drugs in Oregon. *J Subst Abuse Treat*. 2022;141:108849.
19. Victor GA, Strickland JC, Kheibari AZ, Flaherty C. A mixed-methods approach to Understanding overdose risk-management strategies among a nationwide convenience sample. *Int J Drug Policy*. 2020;86:102973.
20. Hathaway AD. Cannabis users' informal rules for managing stigma and risk. *Deviant Behav*. 2004;25(6):559–77.
21. Ridout B, Campbell A. The use of social networking sites in mental health interventions for young people: systematic review. *J Med Internet Res*. 2018;20(12):e12244.
22. Samuel J, Ali GMN, Rahman MM, Esawi E, Samuel Y. Covid-19 public sentiment insights and machine learning for tweets classification. *Information*. 2020;11(6):314.
23. Ali GMN, Rahman MM, Hossain MA, Rahman MS, Paul KC, Thill J-C, et al. editors. Public perceptions of COVID-19 vaccines: policy implications from US Spatiotemporal sentiment analytics. *Healthcare: MDPI*; 2021.
24. Zhong Y, Chen Y-j, Zhou Y, Yin J-J, Gao Y. j. The artificial intelligence large Language models and neuropsychiatry practice and research ethic. *Asian J Psychiatry*. 2023;84:103577.
25. Chen S, Guevara M, Moningi S, Hoebers F, Elhalawani H, Kann BH, et al. The effect of using a large Language model to respond to patient messages. *Lancet Digit Health*. 2024;6(6):e379–81.
26. Birkun AA, Gautam A. Large Language model (LLM)-powered chatbots fail to generate guideline-consistent content on resuscitation and May provide potentially harmful advice. *Prehosp Disaster Med*. 2023;38(6):757–63.
27. Sulkunen P, Babor TF, Cisneros Örnberg J, Egerer M, Hellman M, Livingstone C, et al. Setting limits: gambling, science and public policy—summary of results. *Addiction*. 2021;116(1):32–40.
28. Latha K, Meena K, Pravitha M, Dasgupta M, Chaturvedi S. Effective use of social media platforms for promotion of mental health awareness. *J Educ Health Promotion*. 2020;9.
29. Samuel J, Rahman MM, Ali GMN, Samuel Y, Pelaez A, Chong PHJ, et al. Feeling positive about reopening? New normal scenarios from COVID-19 US reopen sentiment analytics. *IEEE Access*. 2020;8:142173–90.
30. Sirola A, Kaakinen M, Oksanen A. Excessive gambling and online gambling communities. *J Gambl Stud*. 2018;34(4):1313–25.
31. Balem M, Perrot B, Hardouin JB, Thiabaud E, Saillard A, Grall-Bronnec M, et al. Impact of wagering inducements on the gambling behaviors of on-line gamblers: A longitudinal study based on gambling tracking data. *Addiction (Abingdon England)*. 2022;117(4):1020.
32. Trujillo A, Cresci S. Make Reddit great again: assessing community effects of moderation interventions on R/the_donald. *Proc ACM Hum Comput Interact*. 2022;6(CSCW2):1–28.
33. Binde P. Gambling advertising: A critical research review. London: Responsible gambling trust; 2014.

34. Parke A, Harris A, Parke J, Rigbye J, Blaszczyński A. Responsible marketing and advertising in gambling: A critical review. *J Gamb Bus Econ*. 2014;8(3):21–35.
35. Hing N, Russell AM, Gainsbury SM, Nuske E. The public stigma of problem gambling: its nature and relative intensity compared to other health conditions. *J Gamb Stud*. 2016;32:847–64.
36. Hing N, Nuske E, Gainsbury SM, Russell AM. Perceived stigma and self-stigma of problem gambling: perspectives of people with gambling problems. *Int Gamb Stud*. 2016;16(1):31–48.
37. n Der Maas M, Shi J, Elton-Marshall T, Hodgins DC, Sanchez S, Lobo DS, et al. Internet-based interventions for problem gambling: scoping review. *JMIR Mental Health*. 2019;6(1):e9419.
38. Ross AM, Morgan AJ, Jorm AF, Reavley NJ. A systematic review of the impact of media reports of severe mental illness on stigma and discrimination, and interventions that aim to mitigate any adverse impact. *Soc Psychiatry Psychiatr Epidemiol*. 2019;54:11–31.
39. Competiello SK, Bizer GY, Walker DC. The power of social media: stigmatizing content affects perceptions of mental health care. *Social Media + Soc*. 2023;9(4):20563051231207847.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.