Hindawi
Advances in Human-Computer Interaction
Volume 2019, Article ID 9769246, 18 pages
https://doi.org/10.1155/2019/9769246

Research Article

A Risk Management Framework for User-Generated Content on Public Display Systems

Pedro Coutinho [1,2] **and Rui José** [2]

[1] *School of Technology and Management, Polytechnic Institute of Viana do Castelo, Viana do Castelo, Portugal*
[2] *Centro Algoritmi, University of Minho, Guimarães, Portugal*

Correspondence should be addressed to Pedro Coutinho; p.coutinho@estg.ipvc.pt

Received 2 July 2018; Accepted 1 February 2019; Published 3 March 2019

Academic Editor: Hideyuki Nakanishi

Digital public displays can represent a powerful medium for personal expression and situated communication. However, before they can actually serve as an effective communication medium, they need to move towards more open models, in which user-generated content can play a more prominent role in their relevance and value proposition. The key challenge, however, is how to share control with users while being able to guarantee that published content matches the social expectations of a place and the goals of the display owner. In this study, we explore a risk management methodology as a comprehensive approach to this issue. We propose a framework that supports the systematic elicitation of the risks involved, their prioritisation, and the selection of the specific combination of moderation techniques that is able to reduce risk to a level that is deemed acceptable, while minimising the moderation effort and the impact on the willingness of users to publish their content. With this overall framework, we expect to help display owners to reason about the moderation needs of their displays and the best mapping between those needs and various moderation techniques.

1. Introduction

Ubiquitous and mobile technologies open new opportunities for situated digital services that deliver shared locative experiences. In particular, digital public displays are increasingly ubiquitous in urban landscape and they have a unique capability to expose their message in a contextually relevant way to everyone around, breaking personal filter bubbles [1] and enabling rich situated shared experiences. However, current display systems are essentially a world of closed display networks, where only a few people are allowed to post content in narrowcast models. A key enabler for the transformation of public displays into an effective and open communication medium [2] is the ability to allow users to contribute with their own content for the displays. By accepting user-generated content from people in their vicinity, public displays can become truly situated devices, reflecting the contexts in which they are inscribed and the social practices and contexts around them [3]. While this may essentially seem like a benefit for users, empowering users as content creators may also offer important benefits for display owners, e.g. , improving the relevance of the content to their audience, promoting a sense of community, or strengthening the bond with guests. Studies on the practices associated with public notice areas [4, 5] have shown that even though most promoters of community boards had difficulty articulating specific motivation for keeping them, they recognised intangible benefits in the idea of providing a service valued by the community.

Despite widespread acknowledgment of the potential benefits associated with user-generated content, there is also major awareness regarding the fundamental challenge of how to share control with users while being able to guarantee that content published on public displays will stay aligned with the wider social expectations and practices of each place. Unless there is an effective way to frame people's self-expression within the expectations of appropriateness for that place, sooner or later, abusive use will occur. When that happens, any potential value associated with user-generated content will easily be overshadowed by the negative impact of inappropriate content. Consequently, user-generated content on public displays can only be realistically considered within the scope of some type of content moderation strategy.

1.1. The Specific Challenges of Content Moderation on Public Displays. The need for content moderation on public displays is shared with many other types of online communities and social networking platforms. Those platforms are strongly reliant on user-generated content, but poor content curation can easily lead to greater noise, which will then lead to a less useful system and ultimately to its dismissal by most users. This is also an obvious problem for public displays, but despite many similarities, content moderation of user-generated content on public displays is a slightly different type of problem with its own specific challenges.

The first and most striking specificity is that the public nature of public displays makes content moderation particularly sensitive. This is content that will be exposed to whoever is passing-by, which can be potentially very diverse audience, with very different backgrounds, age profiles, and values. At least some of these people are likely to have more sensitive views on what might constitute appropriate content for a public or semipublic place. When faced with what they regard as inappropriate content on a public display, they might feel ambushed by a situation they did not seek. This is very distinct, for example, from a Facebook page, where one may expect to find views that resonate well with the respective audience, even if they could be seen as totally inappropriate by many others.

A second specificity is the high degree of endorsement associated with media posted on public locations [6]. Even when content is user-generated and clearly marked as such, people assume that the owner of the display has somehow approved that content and is therefore endorsing it. They will say that the display at that place was showing certain content, rather than say that a particular person was inappropriately posting certain content to that display. This places additional responsibility on the display owner, who is expected to act as the guardian of appropriate place behaviour.

A final distinction is the physical scale associated with public displays. Content published on a place-based public screen is usually specific to that place and thus limited to a small area where it will only be seen by passers-by. Even in crowded venues, this is always a much smaller scale than what happens in most web services. In a way, this might seem like a benefit because the potential impact of inappropriate content could also be smaller. However, it also means that there is no critical mass for complex moderation techniques. Most display owners are likely to have very little time for content management, and there will not be enough users to support any forms of large-scale crowdsourced moderation.

1.2. Research Objectives. Our research is concerned with the issue of how to open public displays to user-generated content, while mitigating the risks associated with inappropriate content. To address this challenge, we explore an approach inspired by risk management strategies. Risk management [7] is a systematic process to identify, assess, and prioritise risks, so that proper actions can be taken to minimise, monitor, and control the likelihood and/or impact of unfortunate events. In our case, we aim to conduct a systematic elicitation of the risks associated with user-generated content on public displays and analyse the possible role of multiple moderation techniques. While previous work has studied specific moderation techniques for particular risks [8–11], we aim to provide a systematic identification of those risks and techniques. We also aim to address the broad range of control sharing situations, their diverse requirements and the broad range of moderation techniques that can be applied. More than proposing any specific moderation approach, our goal is to offer designers of interactive displays a framework they can use to map their concrete moderation needs to the most suitable set of moderation techniques.

Our methodology combines a qualitative review of previous work on publication paradigms for large screen displays and interviews with 36 potential displays owners. Based on data gathered from these two sources, we propose a comprehensive list of the key risks associated with user-generated content; a prioritisation of those risks according to the perception of potential display owners; a list of the major categories of premoderation and postmoderation techniques; and an overview of the acceptance by place owners of those various moderation techniques. This contribution is a relevant step towards a broader perspective on how to address the risks associated with user-generated content to public displays.

2. Related Work

2.1. Shared Control in Public Displays. Allowing user-generated content on public digital displays is broadly recognised as a key feature for peoples' engagement with the system. The idea of creating displays that reflect the local community in which they are inscribed has been explored from many different perspectives [12], with particular incidence in work environments as a means to disseminate information or provide awareness about group activities [13, 14]. The Funsquare application [15] presents trivia information in a way that reflects the current context around the display. Memarovic et al. [16] present several examples of projects and systems that demonstrate the value of open approaches in engaging passers-by and other actors with the displays and the environment where they are placed. A representative example is the Moment machine [17], where people are allowed to take situated snapshots through a display-attached camera, which are then displayed on an urban screen facing the street. Studies on the privacy implications of posting photos of individuals in the public space [8] have shown the many subtle and diverse ways in which the boundaries between the public and the private spheres can be challenged by these new ways of content publication on public displays. Social media provides a continuous stream of updated content, while preserving the social meaning of the content to the people in the surroundings of the display. Still, Hosio et al. [18] report on how collecting user-generated content from social media platforms can also generate conflicts when that content is shown on public displays. This previous work clearly highlights the continuous tensions between user-generated content and the concerns of appropriateness in public space. These concerns are often expressed in very subtle ways, requiring specific control and moderation strategies.

2.2. Moderation Techniques. The need for moderation and other control mechanisms has already been widely acknowledged in previous research and clearly identified as a requirement for open displays networks [2]. The wide range of public display systems and their particular publication requirements lead to the emergence of many moderations approaches. Greis et al. [19] present a broad study of premoderation techniques, addressing issues related to people's expectations regarding the content moderation process. Regarding a campus deployment, Elhart et al. [20] describe a distributed postmoderation process involving the collaboration of University staff, which allows display owners to keep control over content publication, even when it comes from third-party applications. Elhart et al. [21] suggest that applications need to provide additional information to display owners, based on content's description but also on ratings from other display owners or even display viewers. Taylor et al. [9] study how moderation can be delegated to users that act as trusted curators for a specific content category. Alt et al. [4] study the use of a report abuse functionality in the Digifieds system, which is extensible to the community, allowing the distribution of moderation's effort by a wider set of stakeholders. The Instant Places framework [22] enables people to express their content preferences in the form of keywords in their Bluetooth names. These are recognised when the user checks-in to a display and used to fetch related images from Flickr. The study has shown how some users went through a lot of effort to try to manipulate the publication system in order to push images that were potentially inappropriate.

Social accountability is referred to as the key driver for moderation in the Plasma Network [23], where user authentication and a restricted work environment were enough to prevent inappropriate content from being published. Storz et al. [24] extend this notion of social moderation and suggest the use of social media, not just for bringing content to the displays, but also as a long-term moderation approach. Goncalves et al. [25] suggest a crowdsourced moderation process that encourages the crowd in the surroundings of the display to get involved in moderation activities. The Ubinion service [26] appeals to civil participation of young people to give personalized feedback on municipal issues. They explore users' generated content directly entered in the public display and the use of social media's "liking" and comment facilities for selecting and moderating that content. Results suggest that this kind of service can be valuable to collect feedback from otherwise passive users but also to engage them in community-based moderation.

Publication practices around traditional public notice areas have been studied by Alt et al. [5] as a design inspiration for the emergence of new practices around digital displays. The study addresses the motivations that venue owners may have to share their public boards and also their practices for controlling that content.

Despite these contributions, the current state of the art has not yet provided a systematic framework for approaching the issue of moderation from its many perspectives and help to define the control sharing strategy for a concrete scenario. This work is novel in how it takes this broader perspective on the risks associated with user-generated content and frames current moderation techniques under a broader risk management framework where they can be analysed as an integrated solution to control sharing on a concrete scenario.

2.3. Risk Management. Risk management is a wide topic on its own, with multiple standards, research topics, and a broad set of practitioners across many industries. Organisations of all types and sizes face external and internal factors, which may have a major influence in their ability to achieve their objectives. Risk management is an iterative and systematic process for dealing proactively with those uncertainties and their impact. The ISO 31000 standard [7] provides generic guidelines and terminology for risk management by organisations. This standard is expected to provide a common approach to managing any type of risk, and is not industry or sector specific.

Risk management is also becoming increasingly important within Information Technologies. In particular, software development projects are known to involve many execution risks [27]. The emerging discipline of software risk management [28] attempts to identify, address, and eliminate risk items before they become either threats to successful software operation or major sources of software rework. The use of information technology (IT) in organisations is also subject to various risks [29]. Risk management can play a critical role in protecting an organisation's information assets from such IT-related risks. It is a tool through which IT managers can balance the operational and economic costs of protective measures and define the strategy for protecting the mission-critical IT systems and data [30].

Risk management principles can be applied across many application domains, but they are primarily conceived for the needs of larger projects. This means that existing standards and tools are not a good match to the specific needs of content moderation on public displays. However, the generic principles of risk management provide a consolidate body of knowledge and terminology that can offer the consistency and depth that is needed to approach content moderation as a risk management process.

3. Research Design

Our research design is framed by risk management methodologies, which provide the scaffolding upon which we organised the specific research activities of our work. To formulate content moderation challenges as a risk management problem, we start by defining risk as the effect of uncertainty on objectives. In our specific problem domain, the key stakeholder is a place owner, whose objective is to offer valuable content to its visitors by incentivising people to provide that relevant content themselves, leveraging their effort and possibly their connection to the local community.

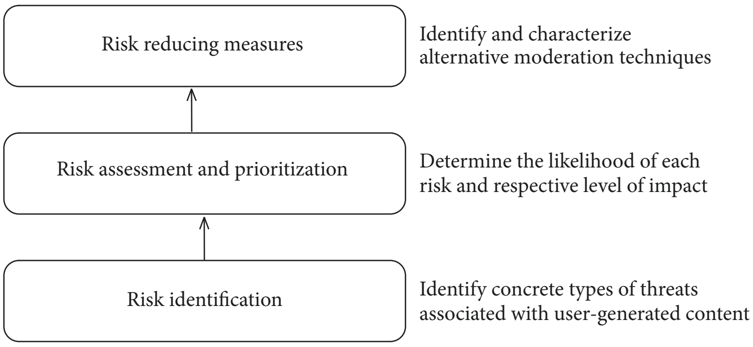

Risk reducing measures
Identify and characterize alternative moderation techniques
Risk assessment and prioritization
Determine the likelihood of each risk and respective level of impact
Risk identification
Identify concrete types of threats associated with user-generated content

Figure 1: Risk management elements for user-generated content.

This objective can be affected by uncertainty regarding the lack of content publishers but also and especially by the possibility of having irrelevant or even inappropriate content on their displays. Visitors are also stakeholders because they can be affected in their objective of having a nice experience at a venue. This may happen if they are confronted with the presentation of inappropriate content or simply annoyed with irrelevant content.

Despite the broad diversity of risk management standards, methodologies, and frameworks, most risk management approaches involve some variation of four fundamental components: risk identification; risk analysis (assessment and prioritisation); risk-reducing measures; and risk monitoring [29]. In our work, we will consider how the first three can be used as a framework for managing the risks of user-generated content on public displays (c.f. Figure 1).

Risk identification involves the characterisation of the potential threats and the assessment of the vulnerability of critical assets to specific threats. Within the scope of our problem domain this mainly involves the identification of concrete types of threats associated with user-generated content. To support this identification, we have conducted a systematic elicitation of the threats associated with user-generated described in previous research.

Risk assessment and prioritisation aim to determine the magnitude of the risk associated with the various threats. This involves determining the expected likelihood of each risk (high, medium, and low) and the respective level of impact (high, medium, and low). The respective magnitude is normally estimated as the product of these two variables. To support the estimation of risk and impact we use data obtained from interviews with place owners.

Risk-reducing measures involve the identification of ways to reduce the risks that have been identified and a prioritisation of those measures based on risk assessment and strategy. In our case, this involves the systematic identification and characterisation of alternative moderation techniques and their assessment. To support the identification of moderation alternatives, we have used our analysis of previous research. To assess their acceptance by place owners, we have used data from place owner interviews.

Our research design was thus determined by the need to obtain grounded data for the various elements of our risk management framework. This was essentially based on two major activities: a qualitative analysis of the literature and interviews with 36 place owners to understand their perspective about potential risks and mitigation strategies. We will now describe these procedures in more detail.

3.1. Qualitative Analysis of the Literature.

A key data source in our research design was a qualitative analysis of moderation situations referred in the research literature. Using specific search tools, such as Google Scholar, Science Direct, Scopus, and Web of Science, we performed an exhaustive literature search. The search criterion was focused on papers with less than 10 years, addressing openness and moderation issues in the context of public displays. In this process, we selected 26 scientific publications (listed in Supplementary Materials (available here)) addressing different facets of this topic. These 26 papers were analysed using a process based on Grounded Theory. We used a coding tool to code any text segments referring to moderation processes, including the different techniques, general concerns, and motivations. The result was a collection of 100 coded segments corresponding to 23 unique codes.

We then conducted a consolidation process, based on an affinity diagramming where we aggregated the various codes into major categories. The result was the identification of 5 top-level concepts, corresponding to the various perspectives of moderation arising from the literature survey, more specifically:

(i) heuristics (concretes experiences on moderation usage);

(ii) inappropriate content (references to various concepts of inappropriate content);

(iii) moderation approaches (references to moderation approaches and techniques);

(iv) moderation evaluation (evaluation of the impact and consequences of moderation in publications' quantity and quality);

(v) motivations (underlying motivations for moderation processes).

This research activity produced two major contributions to our study. The first was a thorough perspective on the various types of risk associated with user-generated content across the research literature. This was the major input for the first part of our risk management approach, which is risk elicitation. The comprehensive view of the risks generated by this process revealed itself to be far more extensive than what could have been possible just from asking directly to places owners. As we have discovered during the interviews, their mindset is strongly biased towards offensive content. The second major contribution of this literature analysis was the identification of a diverse set of moderation techniques. This was the major input for the identification of risk-reducing measures.

3.2. Interviews with Place Owners. The other major data source in our research was obtained through a set of 36 interviews with people responsible for different types of places. The goal was to get a deeper understanding of their perception about the risks of user-generated content and about the moderation techniques they were more willing to use to mitigate those risks.

The first part of the interview was focused on risks. The interviews started with the presentation of 7 key threats emerging from the literature analysis. This presentation was based on the display of content images representing concrete situations associated with each of those risks. The goal was to make sure that participants could easily perceive the concrete nature of the risk and not just some abstract interpretation, allowing them to actually reflect on what these risks were and what they could mean for their own places.

Participants were then asked about their own assessment of the risk (likelihood that a specific risk event could occur) and the respective impact (how harmful or undesirable it would be if it happened), always considering the specific context of their own places. Participants replied using a 5-point Likert scale where they could classify risk situations according to their likelihood (from “unlikely” to “almost certain”) and potential impact (from “minimal” to “maximum”). To avoid any learning effects, the order in which the risks were presented varied from interview to interview. Participants were then asked to give their opinion or make any other comments about those risks.

In the second part of the interview, we aimed to obtain the perception of these place owners about which moderation techniques they would be more willing to use to control user-generated content. Place owners were asked to consider a scenario of a digital display in their own venue, where visitors were allowed to publish their own content. In this scenario, the place owner would always have the capability to remove any content at any time. The set of moderation approaches used in the questions corresponded to those emerging from the literature analysis. Place owners were asked about their acceptance level regarding each of those moderation approaches. Participants replied using a 5-point Likert scale where they could classify their acceptance level (from “Totally unacceptable” to “Totally acceptable”). Participants were also asked to comment on the possible use of each of those approaches in their venues. The interviews were recorded and later transcribed for qualitative analysis. Extracts from those interviews are included in the results as participant citations and provide important complementary insights into their own perspective on these topics.

4. Threats of User-Generated Content

A first contribution of this work and a stepping-stone for our proposed framework is a thorough identification of the key threats resulting from user-generated content on public displays, extending the results from our previous work [31]. Generically, the main threat is the possibility that inappropriate content ends up being shown on the displays. The challenge, however, is to go beyond the diffuse and contextual nature of appropriateness. While most people could quickly point out concrete examples of clearly inappropriate content, they would normally find it very hard to clearly state what exactly distinguishes acceptable from unacceptable content. Humans are very diverse in the relevance they attribute to different values, and this will ultimately lead to conflicting views on what may or not be appropriate content. As a result, moderation is often about place making and seeking the right balance between conflicting views of different members of the community. Humans are also very sensitive to context and the corresponding implications for expected behaviours. In situations of everyday life, appropriateness can thus be highly ambiguous, subjected to many social interpretations, and also very fluid.

A general framework for risk management in user-generated content should take a comprehensive approach to these issues and provide a thorough overview of the many subtle issues involved. To uncover those many and potentially very subtle threats, we conducted a systematic analysis of the research literature on this topic. Using a qualitative analysis methodology, we coded any references to situations where user-generated content was described as a source of harmful or undesirable consequences. We selected those coded segments and recoded them according to the nature of the respective threat. We have only considered the threats where the publishers were using the normal features offered by the service. We have not considered any hacking possibilities or intentional misuse by legitimate system administrators. After a consolidation process, we arrived to a set of 7 risk categories: offensive content; spam; soft hacking; etiquette breach; editorial conflict; copyrighted material; and personal exposure. These categories where then used as the structure for place owner interviews, where we were able to complement the identification of those risks with deeper insights into how they are perceived by those place owners. The final characterisation of these 7 categories highlights the wide range of challenges involved and may help to approach risk management from a broader perspective.

4.1. Offensive Content. The possibility to see offensive content posted on the public display is the most obvious fear associated with user-generated content. Regardless of its specific nature, we consider offensive content as content that most people will perceive as disturbing and clearly inappropriate for most public contexts. Without proper control, sooner or later someone will end up posting explicit material, e.g., adult content, horrible injury, or ostensively aggressive messages. However, even lighter forms of content may in certain contexts be seen as offensive or clearly inappropriate, e.g., swear words or excessively informal language, as participant P6 stated about verbal language used in his venue:

> *"The impact is high because I have clients who speak correctly, but I also have clients who speak swear words and, when they talk, the other people around get disturbed".*

An interesting point is concerned with the attribution of responsibility when a user posts offensive content. In the place owner interviews, it was clear that they believe people would be able to distinguish between their content and offensive content:

> *"People know me well enough and would not put stuff like this here, and if they did, it would not be associated to me" P13;*

> *"The impact would be moderate because people would not associate the content with us and would perceive that it had been placed there without our consent" P24.*

Still, even if they see it as being the result of an obviously malicious and intentional act by a third-party, there is still the issue of the extent to which viewers are going to interpret that publication as a gross failure of the duties of the display owner. Place owners seemed to have mixed views on this topic, as stated by participant P25:

> *"There are always the jokers. . . Belonging the screens to our service, there could be complaints about inappropriate content. . . but people would not associate it with the service, although the responsibility of the content is always of the service".*

This particular threat is unique in how it is so strongly present in people's minds. If we had based the identification of threats solely on interviews with place owners, we would probably not go much beyond this particular risk. The impact on the image of the place can be so negative that avoiding offensive content is normally seen by display owners as their key concern in regard to user-generated content.

4.2. Spam. One of the most recurring problems in social media platforms is spam, which includes more or less obvious forms of advertising. Very often, content being posted as genuine content is actually just a disguised way to promote people, businesses, or content sources, often including branded images with URLs or other contact information. In most cases, spam content will not be perceived as offensive and occasional spam content can even go unnoticed, as stated by P31 concerning the impact of spam in a hypothetical display in his venue:

> *"I believe the impact is moderate because people no longer care much about spam content".*

Still, a system that is not able to handle spam properly can easily see the value of user-generated content being undermined by the noise produced by widespread spam.

Again, a major challenge is how to define the boundaries of what is appropriate and what is spam. A previous study on the distribution of paper leaflets in cafés has shown the diffuse nature of what is acceptable [32]. While place owners can be very sensitive to content from possible competitors or content that could be seen as undifferentiated advertising, they can be very open to specific types of content. For example, event announcements or nearby attractions were regarded as acceptable because they were seen as being relevant to their guests, as referred to by P5 about the relevance of third-party content:

> *"When we present here information from others, people will expect this to be a place where they find interesting information. That would bring more people, those who publish and those who find value on what is published".*

Interestingly, many display owners would even accept content from competitors, as long as it was part of a reciprocal relationship where the competitor would also accept their content. The role that these tacit connections can have in defining what gets accepted shows how these decisions can be highly subjective and strongly embedded with local knowledge.

Spam is already common with nondigital media, but the natural constraints of physical existence mean that publication costs are proportional to the scale of publication. This represents a natural barrier to the scalability of abusive behaviours. Also, with nondigital media, publication occurs in overt mode, where people can be seen posting their content. A shift to a digital medium would break away from physical constraints, significantly exposing displays to more continuous and intensive spam pressure. In a digital environment, publication could potentially occur anywhere at nearly zero cost, and this is one of place owners' concerns. P11 refers to the risks of existing spam campaigns:

> *"Given the aggressiveness of these campaigns (on windshields, on street furniture, on our own shelves). . . if we could have this medium [public display] available, people would take advantage to make this kind of advertising";*

and P24 stated

> *"If there is no security mechanism (control), it will almost certainly happen, especially by outsiders who would see an opportunity here to easily advertise what they wanted".*

The challenge is thus to be able to bring back some sense of locality to the publication opportunities. This would be fundamental to bring some scalability to any control sharing procedures. Therefore, new concepts will be needed to preserve a scalable sense of locality and social relevance. Digital counterparts should strive to introduce some other form of social currency that represents the commitment of publishers. This should enable social negotiation around content, as well as social connections, to remain an integral part of the processes that define the scope of publication and set appropriateness expectations.

4.3. Soft Hacking. In this study, we are not considering the risks associated with security breaches, but there are many forms of hacking that simply try to explore the borderlines of normal system usage to accomplish what may be described as borderline behaviour.

This tends not to be offensive, because the key motivation is the reward for being able to beat the system. It will, however, be something that is provocative enough to show that the frontier is being crossed, something that most place owners would regard as a serious issue, even if it was done just for fun. P11 is in relation to the *likelihood* and impact:

> *"The likelihood is always high because this is an attractive place where many people pass by. The impact is also high because they may be disobeying a regulation [...] even if for the fun of the person who would publish and for the challenge of publishing something for other people to see. This is because people from different age groups come here".*

Several place owners explicitly referred to the possible use of soft hacking as a door for promoting competing places:

> *"Taking advantage of what belongs to others is serious!" P10;*

> *"If it was advertising to a similar business, it could even be more serious." P5;*

> *"If it was another cafe, impact would be maximum." P9.*

This risk is particularly relevant when there are automated moderation procedures involved. For example, in Instant Places [33], users could post words on their Bluetooth names that were then used for selecting images from Flickr. Even though it was difficult to get the system to fetch an image that could be provocative, some users applied considerable time and creativity to the challenge. While this is not necessarily armful, it still needs to be considered, at least to the extent that it is possible to guarantee a robust borderline and avoid opening the door to more serious and ill-intentioned efforts.

4.4. Etiquette Breach. In addition to offensive content and spam, there are many other forms of posting that may be regarded as violating the rules of etiquette for posting in that context. In general, in corporate or institutional environments, any content falling under the category of Not Safe for Work may easily be perceived as an etiquette breach, even if it is content that can easily be found in other less formal public contexts. Whether they are written or not, etiquette rules should be shared and embraced by the community. The concrete ways in which someone may breach the local etiquette will depend on those rules, but common examples include posting offtopic posts or trolling. A troll is a person who publishes deliberately provocative messages to cause trouble, start a contentious topic, derail a discussion, or incite an emotional response from others. This is not necessarily offensive; it is just inappropriate and regarded as undesirable behaviour by the community, as reflected in concerns of participants:

> *"This would be serious because it would generate great discussion, considering the people who come here. I think it would be unlikely anyone would come here and publish, but if they did, we would have a huge discussion. So, the problem would be more for the discussion that it would generate and not necessarily because of the content" P7;*

> *"This is a quiet place and people who come here are peaceful, so I think it is unlikely to happen. The impact is moderate because I do not think content would be associated with us... it would be more for the discussion that could generate" P26.*

This is in line with policies from online forums, which frequently ban topics, such as religion or politics, because of the strong emotional and heated discussions that these topics may generate.

4.5. Editorial Conflict. A particularly subtle threat is when the whole problem comes down to an editorial perspective. An editorial conflict happens when a user posts something that is perfectly acceptable, certainly not abusive, and perhaps even appreciated by the audience, but which somehow fails to meet what the display owner had envisioned as appropriate, as mentioned by participant P11:

> *"There may be a positioning conflict (e.g. Religion). I admit that people next to the display could even assimilate this type of content as normal, but I, as President, do not think that is appropriate".*

Regardless of the specific moderation mechanisms, opening displays to user-generated content is essentially an act of sharing. It means that the display owner is no longer the sole responsible one for thinking of the display content and needs to give some space to other views on what the display should present. This suggests approaches that build strongly on active user participation and high levels of appropriation. If the system is not open enough to offer a compelling value proposition to users, they may not have enough incentives to post their content. However, even when there is a genuine interest in user-generated content, a display owner may still want to maintain some discretionary control on what goes into the displays:

> *"Using content that is generated by others would be nice... but if that would undermine what I had thought for the screen, it would be serious!" P10;*

or participant P26's concern on the corporate image to be preserved:

> *"I think it's unlikely to happen, given people coming here, but I would not like it because it comes out of my interests. I have a concern with the image and it could be damaged with this kind of things".*

For example, in our work with schools [34], these different visions were very clear and there was a continuous tension between the topics that students valued and what school teachers, who had the role of display owners, perceived as valuable for the school context.

These tensions between publishers and display owners can be particularly frustrating for both parties because the whole problem emerges from unspoken implicit rules that even the display owner will have difficulty in stating explicitly. Other than for those situations that may fall under the umbrella of lack of etiquette, most display owners will not be able to specify a priori what their editorial rules are, beyond the basic position that everything should have to be related to the business of the place:

> *"If someone came here to put anything, it would have to be something to do with the business" P20.*

This is in line with the findings on nondigital community boards, where previous research has shown that their creators did not have a predefined profile for those boards. The actual content that composed the boards had emerged from a continuously evolving social negotiation and the interplay of the interests of the board owners and users [4]. Also, a public space is not normally themed to the extent that it can be clearly focused on a topic. For example, in Facebook pages or other online forums, clearly focused editorial line is essential to attract a specific, but very disperse, audience coming from the entire Internet. In a public space, the audience may change very often and can be very diverse. Therefore, a strong focus on a particular theme will be less common. To attract user-generated content, a display owner may have to accept some flexibility in regard to topics that may be of the interest to the display audience, even when they do not correspond to what the display owner had initially envisioned for the public display, as agreed on by participant P18:

> *"What matters is the customer and we have to respect everyone, so the contents are for them".*

It is this inevitably ambiguous and progressive formulation of the editorial line that makes this risk so hard to manage.

4.6. Copyrighted Material. Even if unintentionally, people are likely to post images, branded logos, text, videos, music, or other materials that are protected by copyrights laws. This is probably the case where it can be easier to have objective guidelines about what is appropriate. However, most place owners would not have the competences to identify all forms of copyright infringement and they would not even be able to assess any type of borderline cases, e.g., those related to more subtle concepts, such as fair use. This is probably why they clearly acknowledge the problem but also seem to have a somewhat lazy attitude about it. As stated by participant P15 in regard to the use of copyrighted materials

> *"There are a lot of things that are protected and people use them without knowing it. Even the press uses it without knowing it. Usually that's not a problem..."*

and participant P8 about his own uses of copyrighted materials:

> *"We publish things innocently. I realise for myself, sometimes I pick up cartoon characters for my cakes and we do not even remember we can have problems".*

This almost lazy attitude in regard to the occasional use of copyrighted material can be a particularly strong characteristic in the type of the small venues targeted in our study and is essentially related with nondigital media.

However, they do seem to be aware that the ultimate responsibility lies with them, as mentioned by participant P13 in regard to his responsibility concerns:

> *"It could have a lot of negative impact because it has protected rights and is on my screen".*

In general, we found explicit references to copyright protection in the guidelines for most social media platforms, incentivising users not to include any content which is not original. Also, the specific liabilities associated with potential law infringement need to be considered and users are often warned about the consequences for them in case they fail to comply with copyright rules. Participant P27 makes a reference to the legal aspects of using copyrighted material and he is peremptory:

> *"Impact is maximum because this is illegal!".*

On the other hand, participant P1 states her concern on the consequences to the place:

> *"The impact would be maximum, because if there is an inspection, we would have problems (me and the person who published), and maybe I even would have more problems because the person would not even identify herself on the publication".*

Consequently, while not necessarily a priority, threats associated with copyright infringement induced by user-generated content should be taken seriously. In particular, place owners should, at the very least, have proper mechanisms for dealing with reports of copyright infringement by quickly removing or blocking the respective content.

4.7. Personal Exposure. The final threat emerging from our study is personal exposure. Content involving individuals can be regarded as inappropriate whenever it exposes those individuals without their consent. We have found many references to this risk in online fora, where usage policies often refer that personal attacks or defamatory statements are not acceptable and users should not post content that frames others in a negative light. The concern here is essentially to ask users to focus discussions on the issues and not on people. This is a case where public displays seem to have some interesting specificities. For example, photos of individuals are frequently posted on social networks, very often without explicit permission from the people in the photo and very often with explicit identity markers that make them easy to track. Still, in general, this is increasingly seen as socially acceptable behaviour, as stated by participant P31 about student's publications on social networks:

> *"Students do this every day on Facebook and do not realize it. Very easily they take pictures and make videos of friends and publish them on social networks".*

With public displays, most people have a different perspective. Even completely public data from social networks, such as name and photo, may be perceived as excessive exposure when shown on public displays, as stated by participant P11:

> *"People already find it natural to expose others on social networks. If they realized they can also do it in this public screen, I am quite sure it would happen too. [. . .] It would be very serious because we would be exposing someone on our public screen".*

Previous research by Hosio et al. [18] has shown that many people see a possible conflict when posting to a public display photos with friends in them, even if these photos are already publicly available online. Participant P25 refers to a similar situation at the respective venue:

> *"This has already happened. In fact, it happens with posters that are sent here with photographs, and people are not very careful about it. Here I receive many posters in which I recognize persons in the images, and I strongly believe there are images shown without consent of the people appearing in those photos".*

In the Moment machine [17] people were in general excited about the idea of taking a photo to be presented at the public display, but the authors also report cases where privacy concerns were raised in regard to where and when those photos were being shown. Also, some people were simply not happy about their photos and wanted to have them removed. The authors report on a particular case where a woman contacted the researchers to remove her photo. She did not want to have photos where she was not looking good, especially not in a place she passed by on a regular basis. Throughout the many weeks of the Instant Places [33] deployment, the only occasion where a poster was rejected was when a bar's customers wanted to publish a poster poking fun at other customer. Even though the content was like a joke between friends and would not be seen as offensive by other people, the display owner refused on the grounds that the display content could be placing one of its customers in an uncomfortable situation. Participants have often referred to their concern about this type of content:

> *"The negative impact would be high because they came to complain and I have to answer for what is going on inside my place" P6;*
>
> *"The negative impact would be high because it would always remain the question of who posted it" P25.*

This shows that even when the negative consequences can affect only a single person, place owners are aware of their if responsibility for the published content.

5. Risks Analysis

Building on the characterisation of the main threats, we move to risk analysis. In this phase, we aim to assess the relevance of the various threats, based on their likelihood and their potential impact, and also prioritise them in the overall risk management strategy. To support this process, we gathered data from the interviews with place owners, where participants were asked to classify the various threats according to their likelihood and potential impact using a 5-point Likert scale. The results are listed in Table 1, where we also included the risk relevance as the product of its likelihood and its potential impact.

Based on those results, we were able to build a global view of risk priorities and a risk matrix to help system designers in the definition of appropriate control strategies. The first approach to prioritisation is to simply assume the relevance index obtained from of the product of likelihood and potential impact. The results of this approach are listed in Figure 2.

These results may seem counterintuitive when we consider that offensive content is only ranked as priority number 4, despite being systematically mentioned as the key concern of place owners. However, there are two elements in our study that may help to interpret these results. The first is that some highly impactful risks are also perceived as not very likely or, at least, not very frequent, as summarised by participant P24:

> *"In all these years, I've never seen this type of content here. People can even place content without authorization, but not of this kind."*

Perhaps more interestingly, this also shows how the focus on offensive content can be mainly caused by the lack of awareness about other less obvious types of risks. When confronted with comprehensive lists of threats, place owners might be making a more rational and thorough assessment of risks and a more balanced distribution of their concerns.

To go to a deep deeper into this issue, and better understand the combined effect of likelihood and impact, we have also created a risk matrix, which combines the likelihood associated with a risk with the severity of the respective consequences. A risk matrix is a particularly useful way to analyse risks when the likelihood and potential impact cannot be estimated with accuracy and precision. It provides a simplified perspective of the risk levels and facilitates decision making. The risk matrix represented in Figure 3, represents the same data from Table 1, but this time with the two dimensions separated.

In this diagram, we can clearly observe the emergence of two main groups. The first is composed by the two types of risk that clearly cause a stronger perception of potential

Table 1: Risk analysis data.

Threat	Likelihood	Impact	Risk
Offensive content	2.45	4.20	10.29
Spam	2.88	3.33	9.56
Soft hacking	3.02	3.02	9.15
Etiquette breach	3.08	3.48	10.69
Editorial conflict	2.60	3.05	7.93
Copyrighted material	3.12	3.30	10.31
Personal exposure	3.02	4.38	13.23

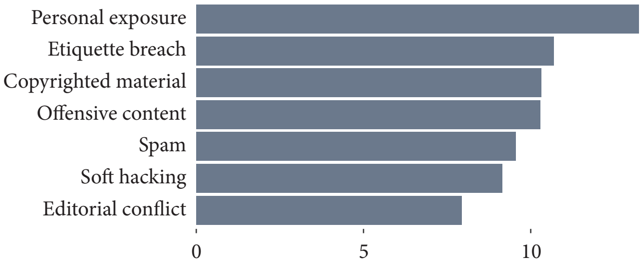


Figure 2: Risk prioritisation based on likelihood and impact.

impact: personal exposure and offensive content. Here, we can highlight how offensive content was regarded as the least likely event from all types of threats. This is why it did not ranked very high in the previous priority list. The second group is a major cluster where all the other threats congregate with similar perceptions of risk and potential impact. Place owners may have had some difficulty in answering with confidence about how likely certain risks would be and that may have lead them to very similar answers closer to safe zone of the centre of the 5 points Likert scale that was used.

5.1. Risk Assessment by Place Type. Another dimension of analysis is to assess the extent to which risk perception changes with the type of places. Figure 4 represents risk perception (likelihood$*$Impact), as assessed by the place owners of the various types of places in our study.

The main observation from these results is the existence of obvious differences between different types of places. Even though we do not have enough places to analyse with more depth the meaning of those differences, we can clearly highlight that risk perception is not uniform and that the social and cultural properties of each place will strongly affect the respective risk perception, as suggested by participant P16 about the impact of offensive content:

> *"Impact would not be high... only men enter here and they would even appreciate it".*

These changes between different types of places are difficult to model, especially because they are likely to be extensive to the different places within the same place type. The key implication is the need for flexible moderation approaches that can easily be adjusted to provide the best possible fit with the unique risk analysis of each place.

6. Moderation Techniques as Risk Reduction Measures

Once risks have been identified and assessed, the next step in a risk management strategy is the identification and selection of risk-reducing measures. From a risk management perspective, risk-reducing measures encompass four major approaches: risk avoidance, risk reduction, risk sharing, and retention [35]. In the context of this work, we are mainly concerned with moderation techniques as risk reduction measures. Here, we extend our previous work on the identification of moderation techniques [36] with the integration of those techniques as part of a comprehensive risk management strategy and the insights from place owners. The elicitation of moderation techniques was based on the literature analysis, where we coded any references to moderation approaches and their properties.

Moderation techniques can themselves be divided into two major groups according to the timing of the process: premoderation and postmoderation. A premoderation approach is a preventive action, where content is moderated before it gets published on the displays, thus reducing the risk of inappropriate publication. Postmoderation represents a set of procedures that can be executed to support moderation after content publication. This is a corrective action that does not prevent inappropriate content from being shown but may reduce the impact caused by situations of inappropriate behaviour. The key advantage of a postmoderation approach is that moderation procedures are no longer an obstacle for a quick publication process, which can be much more rewarding for publishers.

Moderation techniques can also be organised in regard to the entity or entities responsible for the process. Moderation by the display owner can ensure the most effective control but may be hard to scale. Alternatives approaches may involve

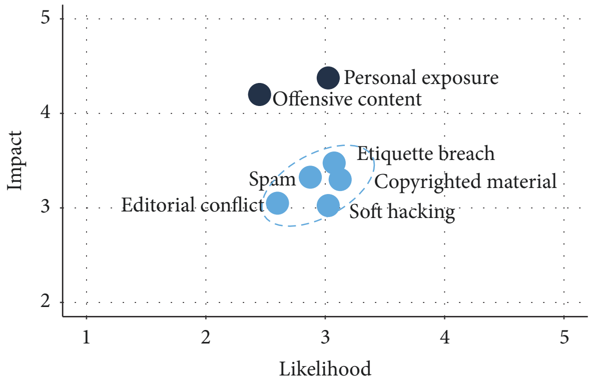


Figure 3: Risk matrix for control sharing in public displays.

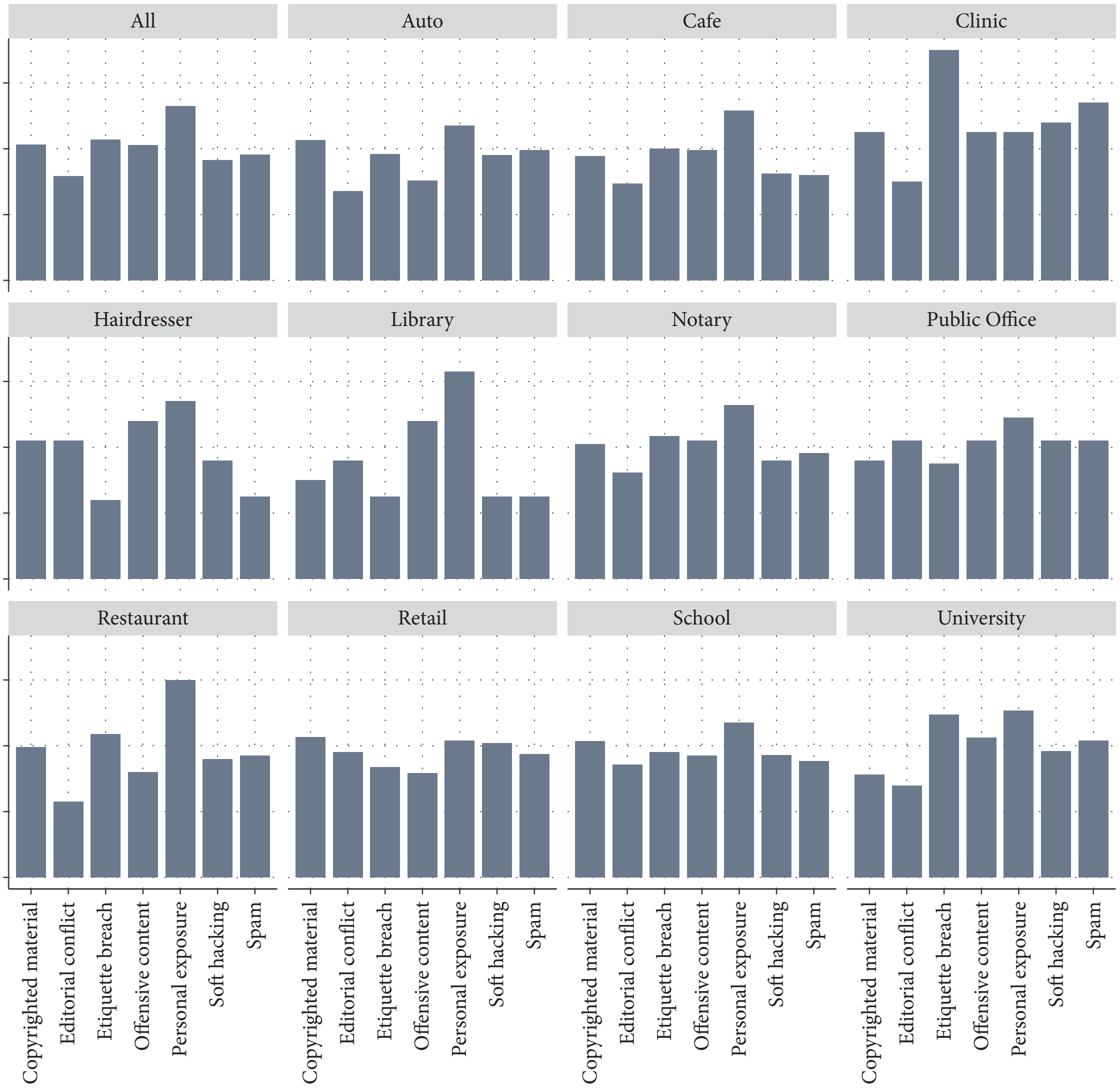


Figure 4: Risk perception by place type.

Table 2: Moderation Techniques.

Actor	Pre-moderation	Post-moderation
Display owner	*Content pre-approval*	*Content reviewing and removal*
System	*Automated filters*	
Trusted curators	*Distributed content curation* *Trusted sources*	*Distributed content removal*
Accountable publishers	*Social accountability*	*Report abusive content*
Anyone		

the automation of the moderation process, the distribution of the process between multiple entities, increasing the accountability of publishers, or selecting who gets access to the right to post content (curation of access). Table 2 summarises the key moderation approaches emerging from the various combinations between moderation timing and moderating entities.

6.1. Premoderation by Display Owner. The most basic form of premoderation involves the preapproval of content by the display owner. A preapproval queue allows display owners to review and approve or reject content before it gets published on the public displays. In their comprehensive study of premoderation techniques, Greis et al. [19] point out that the strong control provided by this form of moderation is a key element for encouraging display owners to publish content generated by others. The key benefit of premoderation seems to be the confidence it can give to display owners about retaining control, as stated by participant P27:

> *"I would open the display to other people as long as I had prior and absolute control over everything they wanted to publish!".*

However, according to Elhart et al. [20] premoderation faces three key challenges: the prior availability of individual content; the scalability of the process; and the negative impact on the publication process due to the publication delays introduced by premoderation. In the interviews, place owners seemed to be aware of many of these challenges, as evidenced by the participant P31:

> *"If we have to control everything before publishing, then there will be no publications. We have this experience and we cannot moderate everything. It is impossible to moderate everything...".*

A large screen display that is open to user-generated content receives a continuous feed of new content posted by users or dynamically fetched from external sources, such as Facebook pages or media feeds. The dynamic nature of this content represents a major challenge for moderation approaches. In closed systems, it is simple to set up approval processes where all contents are carefully screened before publication. In an open system, content is continuously being generated and cannot be known a priori. Any approval mechanisms must be part of regular content management procedures that can keep up with the dynamics of content generation. When content is being generated by applications, the problem can be even more difficult. The exact content that will be shown by an application in a particular contextual situation can be hard to preview until it actually gets presented on the display [21].

Premoderation techniques can also be a problem because of the delays they can introduce in posting/updating content. While in general most users expect content to be moderated, a long delay between posting and having that content on the screen can negatively influence how the users interact with the applications. A study by Greis et al. [19] has shown that delay times caused by content moderation significantly influence the number of user-generated posts on a display. The authors concluded that people accept content moderation on public displays but expect limited publication delays when moderation is done, more specifically within 10 minutes.

6.2. Premoderation through Automated Filters. One approach to reduce moderation delays and deal with highly dynamic content is the use of automated filters that can scan content and place it into quarantine whenever it gets flagged as potentially risky. This is seen as being prone to errors and something that can easily challenge people to soft hacking, as recognised by multiple participants:

> *"There are always things that cannot be filtered and that go through... and if it happens within our space, we are obviously associated with the publication" P17;*
>
> *"I partially agree, as long as the filters are well created, because whoever publishes can always try to escape the filters." P24.*

Despite potential limitations, automated premoderation can be useful without having to be perfect. More specifically, it can complement premoderation by the display owner, improving moderation productivity and the scalability of whole the process. In particular, automated moderation can initially be used only as way to organise moderation tasks. With time and once the place owner realises that direct control is increasingly redundant, he or she may gain enough confidence to allow the system to automate certain decisions and potentially even move to a fully automated process, complemented with postmoderation support.

6.3. Premoderation through Delegated Content Curation. Another way to promote the scalability of moderation processes is to delegate content curation between multiple trusted curators, other than just the display owner. This approach has been extensively explored by Taylor et al. in their study of a Village Photo Display [9]. A group of users was responsible for moderating categories of content created by themselves but approved by the display owner, ensuring that only trusted users could act as moderators. The categories creators were then responsible for the maintenance of the content posted in the category by other users, establishing a form of category ownership. This approach was described as fostering a sense of content ownership by the community and having a positive effect on the number and relevance of the photos to the local community. Also, over the years of this deployment, the authors claim that they were not made aware of any problems with posted content. Participants in our study seemed to appreciate the concept but were not so sure who the trusted curators could be, as stated by participant P5:

> *"That would be ideal, always be seen by someone before going to the screen. But for that, someone should always be available and that is difficult. For example, it could be easier if it were made by people who came here often, who make similar publications..."*

Overall, this distributed content curation approach may also be seen as embracing the use of external sources that are relevant to the place. For example, a Facebook page or an Instagram feed from a trusted source can be seen as curated sources of user-generated content that already incorporate their own content control approaches. The use of social media on public displays provides easy content creation, moderation, and storing, characteristics that can be considered crucial for long-term maintenance of a system [24]. In most cases, the use of content originating from these sources can be considered safe because they already incorporate moderation procedures and their owners have also their own reputation and editorial line to keep. For example, participant P7 would accept content coming from a trusted institution that preserves and even promotes his corporate image:

> *"If I trusted the person... for example, if my game provider told me "let us put it there like that"... if I trusted the external entity coming here, I would have no problem".*

6.4. Premoderation through Social Accountability. Making publishers accountable for what they publish is the other major alternative for premoderation. Even without user authentication, some level of accountability may exist when interaction occurs in overt mode. Previous research has shown that public interaction can generate strong social pressure to the extent that it can even become a huge barrier to the use of public displays for social interaction [37]. This can change significantly with covert interaction, where users are not seen interacting because interaction is mediated by a mobile phone or other similar device [38]. In these cases, user authentication can play a major role. With authenticated users, it becomes possible to define access curation techniques, where only known users are able to post or to make people accountable for their actions:

> *"As long as the person is associated with the publication, that person would always be accountable for her acts,... and she would only do it once, because then she would be banished!" P5.*

Even if publication is open to any authenticated user, this may by itself provide an important barrier for users to post offensive or inappropriate content. It can also help to reduce the perception of endorsement by the place owner, as stated by participant P4:

> *"If their names are there, then it will be clear that this is content that I am not responsible for".*

For example, in their study with tweets, Greis et al. [19] found that forcing people to use a twitter account for posting on a display had a strong impact in the occurrences of inappropriate content. Still, the effectiveness of this approach and the level of trust that is needed can be highly specific to particular communities. In the Plasma Poster Network, Churchill et al. [23] reported on how the restricted physical setting and the relative informality of a workplace were central to the success of the technology and also how a minimal content moderation policy was possible by relying on social accountability to ensure that only appropriate content was posted.

6.5. Postmoderation by Display Owner. The most basic form of postmoderation is to give the display owners simple procedures for the quick removal of any inappropriate content from their displays. In its simplest form, this may correspond to a web page constantly monitored by the display owner that provides a removal option. Whenever new content is brought to the attention of the owner, he or she will have the means to quickly ban that content from the display if it is considered to be inappropriate. This sense of keeping full control over the display is crucial for the willingness to share control with users, as long as the frequency and cost of inappropriate content remains acceptable. This is, however, a corrective action, which does not prevent inappropriate content from being shown, only reduces its potential impact. The key advantage is that moderation is no longer an obstacle for a quick publication process, which can be much more rewarding for publishers.

6.6. Distributed Postmoderation. When used in isolation, postmoderation can be as cumbersome for the display owner as premoderation. There is still the need to frequently monitor new posts to identify and remove inappropriate content. Moderation time is no longer an issue in regard to willingness to publish, but it may affect mitigation of impact. If it takes too long, the negative consequences of presenting inappropriate content may also be too high.

This can be improved by extending postmoderation to trusted reviewers and particularly to people who may be in more direct presence of the displays. Elhart et al. [20] report on a distributed moderation process based on the RFID tags used by people to get access to buildings, which would allow authorized people to interact directly with the display and remove inappropriate content. This process is largely circumstantial, but it avoids the embarrassing and frustrating scenarios of being in front of a display that is presenting inappropriate content and not being able to remove it. It also enables postmoderation to be distributed to a larger number of authorized people, which can make this process much easier to manage.

6.7. Crowdsourced Postmoderation. Postmoderation can also be extended to users by providing a denounce functionality that allows users to report inappropriate content. Since content is already published and being shown to everyone, then everyone can be empowered to denounce content as inappropriate. This is a common approach in crowdsourced platforms, which leverage on the community itself to moderate and define the relevance of the content being shared. When a report occurs, the respective content can be immediately banned and sent to the administrator for verification. For example, in Digifieds [4], users could report inappropriate content through the abuse button. The reported item would be immediately taken out of rotation until reviewed. During the initial six months of deployment, two items with unsuitable content were reported and consequently removed. This possibility to allow everyone to denounce content can make the whole process much more scalable, but it may also have another benefit, which is to allow multiple sensitivities to emerge, highlighting different views on what may constitute inappropriate content and allowing people to express their strong feelings about particular types of content that they find disturbing.

The key problem with this approach is the potential lack of critical mass to make it work. In a media platform, published content may reach a large community in a very short period of time. It will be quickly scrutinized by many who are just a click away from denouncing that content. In a public display, content can potentially be published at a single location, where it may be seen by a few people over a few days. These people may not have any obvious or convenient way to denounce content. Unless it is something very obviously wrong, it may easily stay on the displays for a long time without anyone going through the effort of actually reporting the situation.

7. Acceptance of Moderation Techniques

The final step in our work is to analyse the potential acceptance by place owners of the various moderation techniques. This may be seen as the prioritisation phase of a risk management strategy. However, in this study, the whole process is highly subjective because it depends very heavily on what we see as vague perceptions that place owners might have about a reality that they are only trying to envision. Therefore, to reduce the level of abstraction, we focused the analysis specifically on their perspective about specific moderation techniques. In the final part of the interviews with place owners, we presented them with the various moderation alternatives described in the previous section. The presentation was as specific as possible, with a clear description of the overall approach. We then asked participants to express their availability to operate a system where moderation was solely based on that particular technique. Participants would reply with their level of agreement in a 5 points Likert-type scale, ranging from Total Disagreement to Total Agreement. We have also registered any related comments made during the process.

The first question was focused only on the two major groups of moderation techniques: "Assuming that you could always remove any appropriate content, would you accept to have a display based only on post-moderation techniques?" The answers to this question were overwhelmingly negative:

> *"I had to see everything, whatever it was." P4;*
>
> *"Inside my house I like to see what is going to be published." P21.*

95% of the respondents expressed Total Disagreement, with only 5% going as far as expressing just Disagreement. These results have negatively affected our ability to make any relevant analysis regarding acceptance of postmoderation techniques. We have thus ignored that part of our data and focused only on premoderation techniques. The concrete questions regarding premoderation techniques were as follows:

(i) Preapproval by display owner: I would accept user-generated content if I were able to review any content before it gets published on the displays.

(ii) Automated filters: I would accept user-generated content if there were automated filters, configurable by myself, that would be able to retain most of the inappropriate content.

(iii) Trusted sources: I would accept user-generated content from external sources that I selected as being trustworthy.

(iv) Social accountability: I would accept user-generated content from users who had known identities and could thus be made accountable for their publications.

The results of these questions are represented in Figure 5.

Participants have once more expressed a preference for the preapproval of content by themselves. This is not surprising, considering the context of this study. Participants were not experienced with this type of moderation on public displays and, clearly, they would not be aware of the potential effort associated with a preapproval model centralised on a single person. A few users seemed to be more aware of the implications and clearly mentioned them as a reason to consider other alternatives.

> *"If it was up to me, no one would see anything because I have no time!" P1;*
>
> *"If we have to see everything before publishing, then there would be no publications. We have this experience and we cannot moderate everything. It is impossible to moderate everything…" P31.*

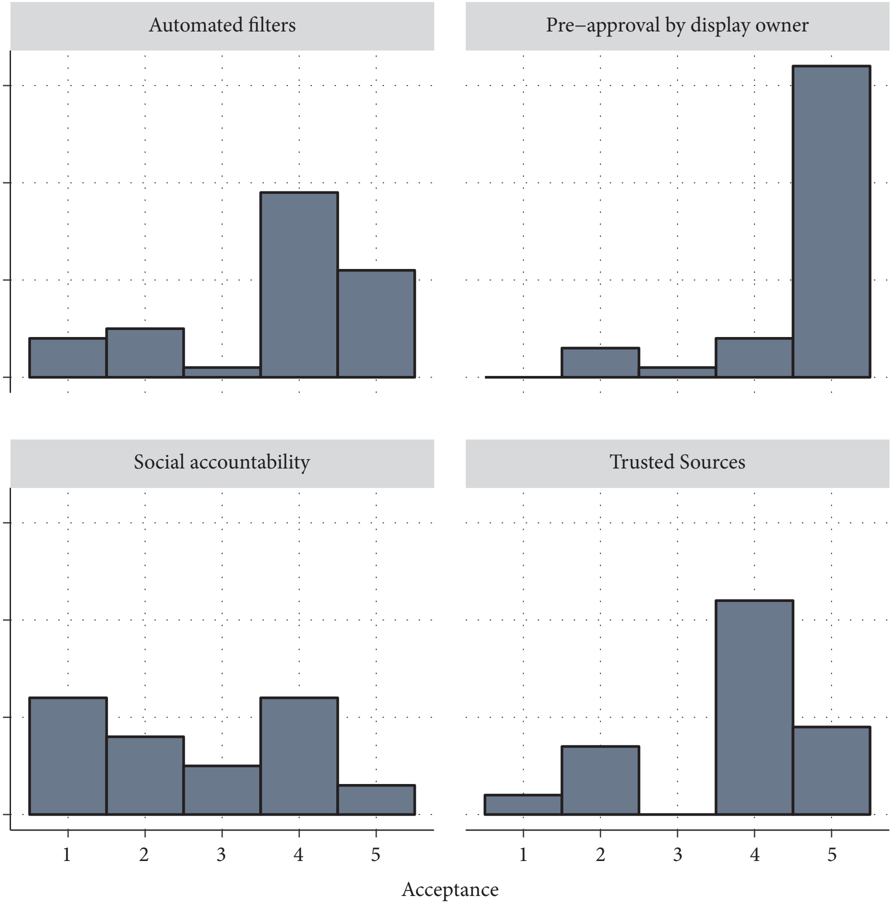


Figure 5: Acceptance of premoderation techniques.

The idea of approving/refusing any user-generated content going to their displays is clearly playing on the safe side. It is also a likely reflection of a focus on the risks of offensive content, as it could actually be a bad solution for other risks, such as copyright infringements.

It is still interesting that the other techniques do not rank so bad. Trusted sources and automated filters seem to raise opposite opinions. This might be related to the preconceptions that participants could have about the effectiveness of those approaches. Social accountability more uniformly spreads across the range of negative and positive opinions. In this case, this might be genuinely associated with the nature of the different communities associated with the various places in this study.

While this may not be good enough to create the expectation that displays would be operated without any type of explicit preapproval, it does show a great potential to the combined use of multiple techniques. For example, automated filters and trusted sources are recognised as approaches to reduce the burden of preapproval by limiting the analysis to borderline cases, as stressed by the participant P6:

> *"I agree partially because the problem is the work it would require. First, I would see and after that would give the OK to publish? The basis here would be trusting on who posted or have some automatic mechanism to filter".*

Social accountability is known to significantly reduce the likelihood of inappropriate content and participants recognise it may help to reduce risks. Still, participants have also mentioned some necessary cautions:

> *"I am afraid things will not be what a person initially thinks they will be" P26;*

> *"Even if the person identifies herself, she can put whatever she likes... we may compromise our corporate image" P24.*

These results suggest the need to combine more than one technique, not just to get better a combination of features but to have the flexibility of adjusting procedures according to evolving circumstances.

8. Conclusions

Enabling users to contribute with their own content can be a huge source of value for communication on public displays. However, user-generated content is clearly perceived as a risky practice, prone to produce abusive appropriations, and uncomfortable situations for display owners and their guests. The obvious uncertainties about the different types of risk and the different techniques that can be used to mitigate them often lead to closed systems or to procedures that demand too much effort from display owners or place too many barriers to user participation. In this study, we have taken a comprehensive approach to the risks of user-generated content on public displays, proposing a general risk management framework for dealing with the various sensitivities of the problem.

The first contribution of this work is the identification of the diverse types of threat associated with user-generated content. People tend to focus only on the high-profile threat of having offensive content on their display. Raising awareness about the full range of risks involved is an important first step towards a general approach to risk mitigation. Previous work has already identified these different risks in regard to specific situations. Our goal was to systematize those different threats into a comprehensive and actionable list of risks.

The second contribution is an assessment of the different types of risks according to the perception of display owners. Based on the results of interviews with 36 display owners, we assessed the perceived likelihood and potential impact of those risks. We were then able to create a global perspective of risk priorities and a risk matrix to help system designers in the definition of appropriate control strategies. It was clear, however, that risk perceptions can change substantially according to different types of place and even different place owners. This means there is no control sharing strategy that can be pointed out as the most adequate for all situations. Risks are very diverse and their relevance can change significantly for different places. Control strategies should thus be flexible and easy to evolve with the likely evolution of risk perception itself.

The third contribution was a categorisation of different moderation techniques. Based on the qualitative analysis of the literature on user-generated content, we have identified a broad range of premoderation and postmoderation techniques, which we have aggregated around major categories. Together, they provide a toolbox for the selection of the specific combination of techniques that can be more suitable for a concrete scenario. When making this analysis, a display owner should seek the combination of techniques that is able to reduce risk to a level that is deemed acceptable, while minimising the moderation effort and the impact on the willingness of users to publish their content. With this overall framework, we expect to help display owners to reason about their moderation needs and the best mapping between those needs and the various alternative moderation techniques.

The final contribution is an analysis of the acceptance by display owners of the various moderation techniques. We have found that most place owners are only ready to rely on premoderation techniques and would not be available to delegate the process to postmoderation approaches. Not surprisingly, explicit control of content approval is the most widely accepted approach, but acceptance levels are also good for other premoderation techniques. This clearly opens the door for hybrid approaches, where different techniques are combined to get the best results in regard to reducing moderation effort, publication barriers, and the risk of inappropriate content.

Overall, this framework should enable moderation to be approached from a broader risk management perspective. A broader perspective means avoiding focus on a single type of risk or in a particular type of moderation technique. It also means understanding that the goal should never be the full elimination of risk. There are other criteria that need to be considered and balanced against the level of risk, such as the moderation load on display owners or the publication barriers faced by publishers. As we have shown, risk perceptions can vary substantially between places and therefore the right balance between all these criteria will always be a local decision and a decision that is likely to evolve over time. A risk management strategy should offer simple and actionable strategies, but it should also offer multiple adjustment approaches that promote alignment between the techniques used and the evolving reality of user-generated content at each place.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work has been supported by FCT—Fundação para a Ciência e Tecnologia—within the Project Scope: UID/CEC/00319/2019.

Supplementary Materials

This is the list of the 26 references used as input in the coding process in which we aim to identify risks and moderation approaches, when opening public displays to publication of user-generated content. *(Supplementary Materials)*

References

[1] M. Foth, M. Tomitsch, L. Forlano, M. H. Haeusler, and C. Satchell, “Citizens breaking out of filter bubbles,” in *Proceedings of the 5th ACM International Symposium*, pp. 140–147, Oulu, Finland, June 2016.

[2] N. Davies, M. Langheinrich, R. José, and A. Schmidt, “Open display networks: a communications medium for the 21st century,” *The Computer Journal*, vol. 45, no. 5, pp. 58–64, 2012.

[3] K. O’Hara, M. Perry, and E. Churchill, *Introduction to Public and Situated Displays*, K. O’Hara, M. Perry, and E. Churchill, Eds., Kluwer Academic Publishers, Dordrecht, Netherlands, 2003.

[4] F. Alt, T. Kubitza, D. Bial et al., “Digifieds: insights into deploying digital public notice areas in the wild,” in *Proceedings*

of the 10th International Conference on Mobile and Ubiquitous Multimedia, pp. 165–174, Beijing, China, December 2011.

[5] F. Alt, N. Memarovic, I. Elhart, D. Bial, and A. Schmidt, "Designing shared public display networks: implications from today's paper-based notice areas," in *Proceedings of the 9th International Conference on Pervasive Computing*, pp. 258–275, 2011.

[6] P. Coutinho and R. José, "Design sensitivities from public expression practices with non-digital displays," in *Proceedings of the 4th International Symposium on Pervasive Displays (PerDis'15)*, pp. 139–145, Saarbruecken, Germany, June 2015.

[7] A. Dali and C. Lajtha, "ISO 31000 Risk Management— "The Gold Standard"," *EDPACS*, vol. 45, no. 5, pp. 1–8, 2012.

[8] N. Memarovic, "Public photos, private concerns - uncovering privacy concerns of user generated content created through networked public displays," in *Proceedings of the 4th International Symposium on Pervasive Displays (PerDis'15)*, pp. 171–177, Saarbruecken, Germany, June 2015.

[9] N. Taylor, K. Cheverst, D. Fitton, N. J. Race, M. Rouncefield, and C. Graham, "Probing communities: study of a village photo display," in *Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces (OzCHI'07)*, pp. 17–24, Adelaide, Australia, November 2007.

[10] F. Alt, N. Memarovic, M. Greis, and N. Henze, "UniDisplay — A research prototype to investigate expectations towards public display applications," in *Proceedings of the 2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS)*, pp. 519–524, Budapest, Hungary, March 2014.

[11] A. Melro, B. Silva, and R. José, "Media sharing in situated displays: service design lessons from existing practices with paper leaflets," in *Exploring Services Science*, vol. 143 of *Lecture Notes in Business Information Processing*, pp. 322–328, Springer, Berlin, Germany, 2013.

[12] N. Taylor, M. Rouncefield, K. Cheverst, and S. Izadi, "Encouraging community spirit with situated displays," in *Proceedings of the AISB 2008 Symposium on Persuasive Technology*, pp. 39–42, April 2008.

[13] S. Greenberg and M. Rounding, "The notification collage: posting information to public and personal displays," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*, vol. 3, pp. 514–521, Seattle, Wash, USA, 2001.

[14] J. McCarthy, T. Costa, and E. Liongosari, "UniCast, OutCast & GroupCast: an exploration of new interaction paradigms for ubiquitous, peripheral displays," in *Proceedings of the Distributed and Disappearing User Interfaces in Ubiquitous Computing - Workshop at CHI2001*, Seattle, Wash, USA, 2001.

[15] N. Memarovic, I. Elhart, and M. Langheinrich, "FunSquare: first experiences with autopoiesic content," in *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia (MUM '11)*, pp. 175–184, ACM Press, December 2011.

[16] N. Memarovic, M. Langheinrich, K. Cheverst, N. Taylor, and F. Alt, "P-LAYERS — a layered framework addressing the multifaceted issues facing community-supporting public display deployments," *ACM Transactions on Computer-Human Interactions (TOCHI)*, vol. 20, no. 3, pp. 1–34, 2013.

[17] N. Memarovic, A. F. G. Schieck, H. M. Schnädelbach, E. Kostopoulou, S. North, and L. Ye, "Capture the moment - "in the wild" longitudinal case study of situated snapshots captured through an urban screen in a community setting," in *Proceedings of the 18th Conference on Computer Supported Cooperative Work & Social Computing (CSCW'15)*, pp. 242–253, Vancouver, BC, Canada, March 2015.

[18] S. Hosio, H. Kukka, and J. Riekki, "Social surroundings: bridging the virtual and physical divide," *IEEE MultiMedia*, vol. 17, no. 2, pp. 26–33, 2010.

[19] M. Greis, F. Alt, N. Henze, and N. Memarovic, "I can wait a minute: uncovering the optimal delay time for pre-moderated user-generated content on public displays," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI'14)*, pp. 1435–1438, Toronto, Ontario, Canada, April 2014.

[20] I. Elhart, N. Memarovic, M. Langheinrich, and E. Rubegni, "Control and scheduling interface for public displays," in *Adjunct Proceedings of the International Conference on Pervasive and Ubiquitous Computing (UbiComp'13)*, pp. 51–54, Zurich, Switzerland, September 2013.

[21] I. Elhart, M. Langheinrich, N. Davies, and R. Jose, "Key challenges in application and content scheduling for open pervasive display networks," in *Proceedings of the 2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops 2013)*, pp. 393–396, San Diego, CA, USA, March 2013.

[22] R. José, N. Otero, S. Izadi, and R. Harper, "Instant places: bluetooth presence and naming as enablers for situated interaction and user - generated content in public displays," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 52–57, 2008.

[23] E. F. Churchill, L. Nelson, L. Denoue, P. Murphy, J. I. Helfman, and J. Helfma, "The plasma poster network: social hypermedia on public display," in *Public and Situated Displays. Social and Interactional Aspects of Shared Display Technologies*, K. O'Hara, Ed., pp. 233–260, Kluwer Academic Publishers, London, UK, 2003.

[24] O. Storz, A. Friday, N. Davies, J. Finney, C. Sas, and J. G. Sheridan, "Public ubiquitous computing systems: Lessons from the e-Campus display deployments," *IEEE Pervasive Computing*, vol. 5, no. 3, pp. 40–47, 2006.

[25] J. Goncalves, S. Hosio, D. Ferreira, and V. Kostakos, "Game of words: tagging places through crowdsourcing on public displays," in *Proceedings of the 2014 Conference on Designing Interactive Systems - DIS '14*, pp. 705–714, Vancouver, Canada, June 2014.

[26] S. Hosio, V. Kostakos, H. Kukka, M. Jurmu, J. Riekki, and T. Ojala, "From school food to skate parks in a few clicks: using public displays to bootstrap civic engagement of the young," in *Pervasive Computing*, vol. 7319 of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 425–442, Springer, Berlin Heidelberg, Germany, 2012.

[27] B. Flyvbjerg and A. Budzier, "Why your IT project may be riskier than you think," *Harvard Business Review*, vol. 89, no. 9, pp. 601–603, 2013.

[28] B. W. B. Boehm, "Software risk management: principles and practices," *IEEE Software*, vol. 8, no. 1, pp. 32–41, 1991.

[29] K. Bandyopadhyay, P. P. Mykytyn, and K. Mykytyn, "A framework for integrated risk management in information technology," *Management Decision*, vol. 37, no. 5, pp. 437–445, 1999.

[30] G. Stoneburner, A. Gougen, and A. Feringa, *NIST SP 800-30 - Risk Management Guide for Information Technology Systems*, Computer Security Division, 2012.

[31] P. Coutinho and R. José, "Risk elicitation for user-generated content in situated interaction," in *Proceedings of the Ubiquitous Computing and Ambient Intelligence: 10th International Conference, UCAmI 2016, San Bartolomé de Tirajana, Gran Canaria,*

Spain, C. R. García, P. Caballero-Gil, M. Burmester, and A. Quesada-Arencibia, Eds., vol. 10069 of *Lecture Notes in Computer Science*, pp. 481–486, Springer International Publishing, Cham, Switzerland, December 2016.

[32] R. José, H. Pinto, B. Silva, and A. Melro, “Pins and posters: paradigms for content publication on situated displays,” *IEEE Computer Graphics and Applications*, vol. 33, no. 2, pp. 64–72, 2013.

[33] R. José, N. Otero, S. Izadi, and R. Harper, “Instant places: using bluetooth for situated interaction in public displays,” *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 52–57, 2008.

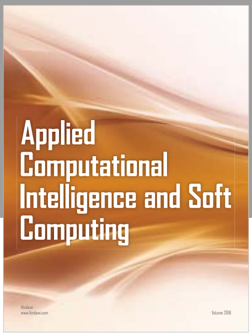
[34] N. Otero, R. José, and B. Silva, “Interactive public digital displays: investigating its use in a high school context,” in *Proceedings of the Move to Meaningful Internet Systems: OTM 2012 Workshops*, vol. 7567 of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 617–626, Springer, 2012.

[35] M. S. Dorfman and D. A. Cather, *Introduction to Risk Management and Insurance*, Pearson, 10th edition, 2012.

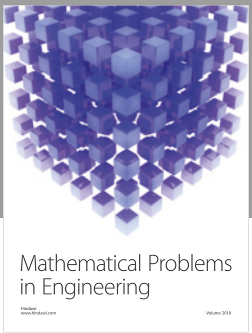
[36] P. Coutinho and R. José, “Moderation techniques for user-generated content in place-based communication,” in *Proceedings of the 2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–6, Lisbon, Portugal, June 2017.

[37] Y. Rogers and H. Brignull, “Subtle ice-breaking: encouraging socializing and interaction around a large public display,” in *Proceedings of the CSCW02 Workshop*, pp. 1–6, 2002.

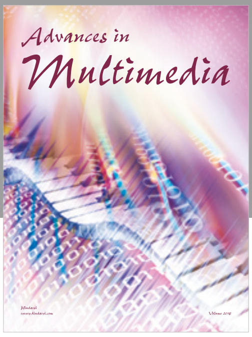
[38] M. Finke, A. Tang, R. Leung, and M. Blackstock, “Lessons learned: Game design for large public displays,” in *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts, DIMEA 2008*, pp. 26–33, New York, NY, USA, September 2008.

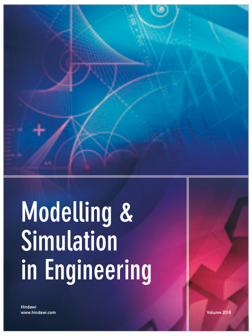

Applied Computational Intelligence and Soft Computing

The Scientific World Journal


Mathematical Problems in Engineering

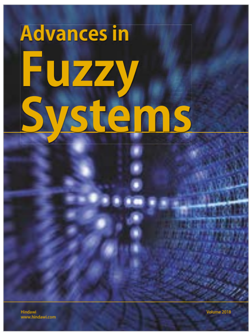
Journal of Engineering

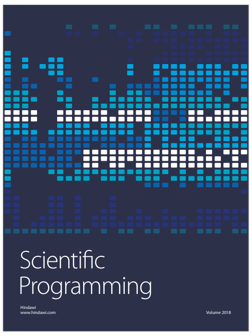

Advances in Multimedia

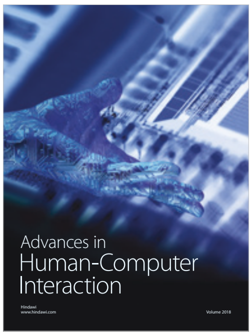

Modelling & Simulation in Engineering

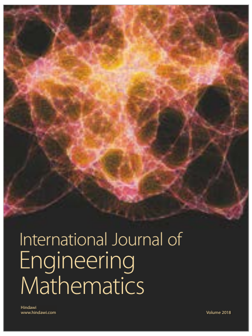
Advances in Artificial Intelligence

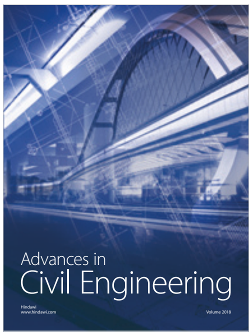
International Journal of Reconfigurable Computing

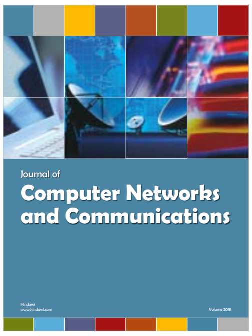

Advances in Fuzzy Systems


Scientific Programming


Advances in Human-Computer Interaction

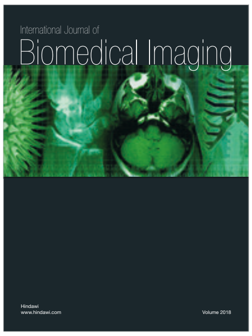

International Journal of Engineering Mathematics


Advances in Civil Engineering

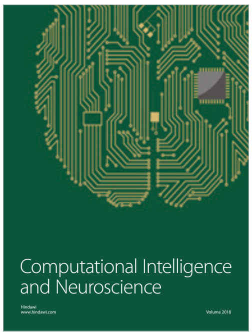

Journal of Computer Networks and Communications

Journal of Robotics

International Journal of Computer Games Technology


International Journal of Biomedical Imaging

Journal of Electrical and Computer Engineering


Computational Intelligence and Neuroscience