Content Analysis and Recommendation Generation

Technical Field

[0001] This disclosure generally relates to methods and systems for the analysis of content, including text content and the generation of recommendations or suggestions based on the analysis.

Background

[0002] This background description is provided for the purpose of generally presenting the context of the disclosure. Contents of this background section are neither expressly nor impliedly admitted as prior art against the present disclosure.

[0003] Communication platforms bring individuals unfamiliar with each other into a common virtual environment where communication occurs. Such communication may occur in the form of text messages, images, video or other formats. The virtual environments may include platforms for the provision of goods or services such as online marketplaces, ridesharing platforms, food delivery platforms etc. Communications may occur between merchants and consumers in various formats such as message threads or forums. Communications may also include reviews generated by consumers and directed towards providers of goods or services.

[0004] Operators of communication platforms or virtual environments such as marketplaces put in place policies or rules that govern communications and rules that users of the platform are subject to. For example, a communication platform may have policies in place regarding the use of offensive language, the confidentiality of personal data, the non-solicitation of service outside a predefined scope etc. In a marketplace, the interests of consumers and merchants may not align in all instances. For example, providers of ridesharing services may wish to minimize waiting time, while users of ridesharing services may seek greater flexibility. The existence of competing interests may result in abrasive communication between participants that may ultimately result in communication that is non-compliant with policies put in place by the operators of the virtual environment. Further, as norms and rules around communication, in particular communication in virtual environments are continuously evolving, evaluating compliance with standards remains a significant challenge. The norms and rules vary from country to

country and culture to culture. In addition, the volume of content generated in virtual platforms has increased exponentially making content analysis a significant computational problem.

[0005] Non-compliance with policies may result in the need for intervention by human moderators, removal of content, warnings to users or termination of a user's access to a platform. Steps such as removal of content may require intervention by moderators which may be delayed because of limited moderation resources. In addition, as content compliance standards change, not all users may be familiar with such changes which may further result in the promulgation of non-compliant content on a platform. It is desirable to provide alternative systems and methods of content evaluation that are less labor-intensive and more efficient at handling undesirable communication in virtual environments.

Summary

[0006] <To be completed after the claims are approved>.

Brief Description of the Drawings

[0007] Some embodiments of systems and methods for content analysis and recommendation generation, in accordance with the present disclosure, are described, with reference to the non-limiting example illustrated in the accompanying drawings in which:

[0008] Figure 1 illustrates a system for content analysis and recommendation generation;

[0009] Figure 2 illustrates a flowchart of a method of content analysis and recommendation generation; and

[0010] Figure 3 illustrates a schematic diagram of some steps and components for content analysis and recommendation generation.

Detailed Description

[0011] Systems and methods of the disclosure perform analysis of content generated in a virtual environment where users are subject to a plurality of content compliance standards. The analysis of content comprises analysis to evaluate compliance of the

content to a plurality of content compliance standards. The embodiments include methods and systems for evaluating compliance of target content as it is generated by a user. Embodiments also generate recommendations for the user in the event the target content is deemed to be non-compliant with the prevailing standards. The embodiments advantageously operate as a moderation bot attempting to deescalate an abrasive communication thread by generating suggestions to users in response to non-compliant content before the non-compliant content is communicated to its intended recipient. The suggestions may include suggestions of alternative compliant content that the user may readily substitute the target non-compliant content with.

[0012] Some embodiments may perform analysis of text content as a user types and may proactively suggest a change to the content, or flag non-compliance and generate a compliant alternative text in a suggestion. Some embodiments may perform the analysis of text content and generate a recommendation after a user submits a send command through their device. The suggestions serve as 'off-ramps' for a communication thread that may otherwise escalate and result in the generation of non-compliant content and poor user experience. The content analysis and recommendation generation occurs in real-time or near real-time with respect to the point of time the user makes the target content available on their computing device for analysis.

[0013] Figure 1 illustrates a block diagram of a system for content analysis and recommendation generation. The exemplary framework of Figure 1 comprises a first remote system 110 in control of user 1 and a second remote system 130 in control of user 2. The first and second remote systems are illustrated for convenience and to simplify the description. In practice, embodiments of the disclosure may be deployed in a scalable manner to a communication platform comprising a large number of users communicating using a large number of computing devices.

[0014] Some embodiments may comprise a separate content analysis system 120 that performs the analysis of content and generate recommendations. In such embodiments, content analysis system 120 may have access to content generated by systems 110 and 130. The content analysis system 120 may direct suggestions to the remote systems 110 and 130. Alternatively, the content analysis and recommendation generation may be performed by the respective remote systems locally. Any suggestions generated may be directly presented to the user of the remote systems.

[0015] The remote systems 110, 130 may be end-user computing devices such as a smartphone, a tablet device, a personal computer etc. The remote systems 110, 130 comprise at least one processor 111, memory 112 comprising program code, one or more input device 115 and one or more output device 116. The input device 115 may include one or more of a keyboard (including an on screen keyboard), a pointing device, a camera, a microphone etc. The output device 116 may include a display, a speaker etc. In embodiments where the remote system 110 is a smartphone, a user may enter text content through an onscreen keyboard on the smartphone. Any suggestions generated by the system may be presented on a display of the smartphone. The various devices are configured to communicate with each other over a network 140 that may comprise a data communication network including the internet or a cellular communication network etc. Memory 112 comprises executable code implementing a communication application 117 that provides the user with the ability to transmit and receive content to other users of the communication platform. In some embodiments, the communication platform may be a part of a platform for providing goods and services such as a ridesharing platform, a food delivery platform or a virtual marketplace platform.

[0016] The communication application 117 comprises a content classification model 113 and a suggestion generation model 114. The content classification model is configured to process a target content and generate a classification with respect to compliance or non-compliance of the target content with a plurality of content compliance standards or policies. The content classification model may be implemented as a supervised machine learning model trained using a training dataset including labelled compliant and non-compliant content. The content classification model may be capable of processing text content, audio content, image content or video content. In some embodiments, in addition to the classification of the target content, the content classification model may also generate a vector representation of the target content and any related content or context information which is further described with reference to Figure 3.

[0017] Text content may be classified in its raw form – i.e. as received. Alternatively, text may be normalized. Text normalization refers to the transferal of text from its original form into a form that can be classified – e.g. a uniform lexicon. This ensures, for example, misspellings, intentionally malicious obfuscations (e.g. the original text may include "sm4ll" instead of the intended word "small") and other anomalies are transferred into text content that can be classified. Normalization can be performed using an adversarial spell

checker or by substitution of known interchangeable characters – e.g. "4" for "a" or "A", and "3" for "e" or "E ". The text content, or normalized text content, is classified as set out above. In some embodiments, classification may leverage a language model or multilingual language model for intent detection and interpretation in text content. IN some embodiments, the multilingual XML-RoBERTa model which can be trained to detect behaviours or intentions reflected in the text content, such as sexual harassment.

[0018] The content analysis system 120 comprises at least one processor 121, a memory 122 that comprises executable code to implement and execute a content classification model 123 (equivalent to content classification model 113) and a suggestion generation model 124 (equivalent to suggestion generation model 114).

[0019] The suggestion generation model 124 may leverage a natural language processing model to generate, from non-compliant language (i.e. language that does not comply with relevant policies) reflected in the text content, suggested content comprising compliant language. The compliant language shall deliver the same or similar message as the non-compliant language, while removing intention or nuance that does not comply with the relevant policies. In some embodiments, a seq2seq model, such as seq2seq T5, may be tuned to translate non-compliant language into compliant language. Moreover, the text content may be normalized to make the text content consistent with language that can be interpreted by the suggestion generation model.

[0020] Figure 2 illustrates a flowchart of a method of content analysis and suggestion generation. At step 210, system 110 (or system 120) receives a target content for analysis. In some embodiments, step 210 may be performed after a user types in a single word or after the user types in a single character or after the user types a phrase or a sentence. The frequency of the execution of step 210 may be varied to optimize the usage of resources on system 110 (or system 120) while proving a good user experience over communication application 117. In some embodiments, step 210 may be executed after a user provides a send instruction in relation to the target content.

[0021] At step 220, compliance of the target content 310 is evaluated by the content classification model 113 (or the content classification model 123). The content classification model comprises program code that embodies computational rules or logic corresponding to content compliance standards 313 set by the platform operator. With reference to Figure 3, the content classification model in addition to the target content 310 may also receive as input the content thread 311. Content thread or series of content

311 may comprise historical communications between user 1 and user 2 to whom the target content 310 is directed. In other embodiments, the content thread 311 may also comprise other content generated by users 1 and 2 that may not be directly related to the target content. The other content may include content generated by users 1 or 2 in their communication with other users. The communication thread 311 may be represented as a vector input to the content classification model and serves to provide context information to improve classification outcomes.

[0022] In some embodiments, the target content may be related to a transaction such as a ride-sharing transaction, a food delivery transaction etc. To improve the classification outcome, the classification model of some embodiments may also incorporate transaction parameters 312 as part of the input to the classification model. The transaction parameters for a ridesharing transaction may include parameters related to service outcomes such as a delay in the arrival of a ride, delay in the arrival of a passenger etc. All of the input to the classification model may be transformed into vector representations before being processed by the classification model.

[0023] Step 230 is performed on determining non-compliance at step 220. In some embodiments, the content classification model may generate an escalation vector 314. The escalation vector 314 may indicate a degree of escalation and/or the nature of non-compliant content. In other words, the escalation vector 314 may encode the information in the target content (and optionally information in the content thread and transaction parameters) in a manner that could be processed by the suggestion generation model 114 at step 230. The suggestion generation model may be implemented using a deep generative model or a generative adversarial network. In some embodiments, the suggestion generation model may be implemented using a supervised classification model that maps an escalation vector to a library of predefined suggestions that serve as labels. In some embodiments, the suggestion generation model may be implemented using customized variations of commercially available text generation frameworks or large language models such as BART or GPT etc.

[0024] The suggestion generation model generates suggestion content that directs the creator of the target content to deescalate their interaction. The suggestion content may also include alternative compliant content 315 that the user may readily incorporate in their communication. By providing alternative compliant content or a suggestion to reconsider the target content, the embodiments advantageously reduce the likelihood of

generation and transmission of non-compliant content on the communication platform. The reduction in transmission and generation of non-compliant content simplifies content moderation tasks for human content moderators.

[0025] At step 240, the generated suggestion is communicated to the user. In embodiments where the suggestion generation is performed on system 110, the generated suggestion is displayed through the output device 116. In embodiments where the suggestion generation is performed by the content analysis system 120, the generated suggestion is first transmitted to the system 110 and it is subsequently displayed by the output device 116. The displayed suggestion in some embodiments may include alternative compliant content and a suggestion to the user to replace the target message with the alternative compliant content.

[0026] Some embodiments may execute the method of Figure 2 iteratively after typing of each character, word, phrase or sentence by a user. In some embodiments, wherein the target message comprises content in different formats, for example text and image content, the steps of the method of Figure 2 may be performed in parallel in relation to each content type and compliance of each content type may be evaluated separately. The content classification model of some embodiments may be an agglomeration of a plurality of models, each model trained to evaluate compliance of a distinctive content type (for example a text content classification model, an image content classification model, an audio content classification model, a video content classification model etc.). Results of each of the models may collectively determine the outcome at step 220 and drive generation of the suggestion content 230.

[0027] The following section illustrates some examples moderation performed by content analysis and suggestion generation features of some embodiments.

<u>Scenario 1</u>

(msg-1) Passenger: *hey buddy, how was your day*

(msg-2) Driver: *good, how are you*

(msg-3) Passenger: *great. Do you want to go to my home*

(msg-4) Passenger: *I will pay you extra 100*

(msg-5) Passenger: *If we can take shower together* ← "policy violation by passenger, soliciting sexual service"

The target message 3 may be processed by the content classification model and be flagged as potentially non-compliant content. The flagging of potentially non-compliant content results in the generation of a suggestion text as illustrated below.

(typed msg-3, but unsent) *Passenger: great. Do you want to go to my home*

(msg-3a, suggestion generated by the suggestion generation model) *Moderator bot: Hi, it is not a good idea to invite a driver to your home.*

Did you mean: "great. Can you drop me off at the door and not at the end of my driveway?"

(msg -3b) Passenger: *Yes*.

In the above sequence, Did you mean: "great. Can you drop me off at the door and not at the end of my driveway?" is an example of alternative compliant content.

If the user ignores generated suggestion and continues:

(msg-4, unsent) *Passenger: I will pay you extra 100*

(msg-4a) *Moderator bot: Hi, all payments must go through an approved platform, including tips. Solicitation for a service that is not allowed could be a problem. The driver may report you to authorities. Or the platform may have to stop from using its services. Anything we can help you with?*

Scenario 2

(msg-1) *Driver: Hi Jen, how's your day*

(msg-2) *Female Passenger: Hello*

(msg-3) *Driver: You look so pretty*

Passenger does not reply. A little later during the ride

(msg-4) *Driver: Are you married?*

Passenger stays silent

(msg-5) *Driver: Do you have a boyfriend?*

Communication thread with the content analysis and suggestion generation

(msg-3) *Driver: You look so pretty*

(msg-3a, suggestion generated by the suggestion generation model) *Moderator bot: This message may make your passenger uncomfortable. They may rate you poorly or even report you.*

Assuming driver ignores and proceeds with his intent

(msg-4) *Driver: Are you married?*

(msg-4a, suggestion generated by the suggestion generation model) *Moderator bot: Asking unnecessary personal information violates the platform's policy. It's unclear why you need this information. Do you want to talk to someone else, perhaps they can help you?*

Scenario 3

(food review) *Consumer: the rate quoted to me was $13 but your f**king malay driver lie and deceit me.*

Consumer submits review. It gets removed later following user complaints.

With content review and suggestion generation in place.

(food review) *Consumer: the rate quoted to me was $13 but your f**king malay driver lie and deceit me.*

Moderator bot generated suggestion content: *Your review may contain offensive remarks that violates Grab's policy for code of conduct. Your account may be suspended for infringing the policy.*

Do you want to talk to our support team about your experience?

Scenario 4

(food review) *Consumer: tastes like licking my toes*

Consumer submits the review. Merchant disputes the review later as ‘derogatory and hateful’.

With content review and suggestion generation in place.

(food review) *Consumer: tastes like licking my toes*

Moderator bot generated suggestion content: *Hi, you seem to have had a poor experience. Descriptive reviews help merchants improve and help other eaters make a decision. Do you want to add more details?*

[0028] As illustrated in the examples above, evaluation for potential non-compliance of target content before it becomes part of a public record or is transmitted to an intended recipient is advantageous in improving the congeniality of communications in a platform. The intervention by the moderation bot of the embodiments reduces the burden on human moderators and directs users away from undesirable conduct where feasible.

[0029] The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavor to which this specification relates.

[0030] Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

[0031] The scope of this disclosure encompasses all changes, substitutions, variations, alterations, and modifications to the example embodiments described or illustrated herein that a person having ordinary skill in the art would comprehend. The scope of this disclosure is not limited to the example embodiments described or illustrated herein. Moreover, although this disclosure describes and illustrates respective embodiments herein as including particular components, elements, feature, functions, operations, or steps, any of these embodiments may include any combination or permutation of any of the components, elements, features, functions, operations, or steps described or illustrated anywhere herein that a person having ordinary skill in the art

would comprehend. Although this disclosure describes or illustrates particular embodiments as providing particular advantages, particular embodiments may provide none, some, or all of these advantages.

Claims

1. A system comprising:

   one or more processors (processor(s));

   an input unit;

   an output unit;

   a memory comprising instructions that when executed by the one or more processors cause the system to:

      receive a target content from the input device;

      process the target content to evaluate compliance with a plurality of content compliance standards;

      on determining non-compliance of the target content, generate a suggestion content in real-time or near real-time; and

      communicate the suggestion content to a user of the system using the output unit.

2. The system of claim 1, wherein the suggestion comprises an alternative compliant content generated based on the target content.

3. The system of claim 1, wherein the compliance of the target content is evaluated based on a series of content related to the target content, the series of content comprising content in a communication thread comprising the target content.

4. The system of claim 3, wherein the target content is text content; and

   the target content is evaluated and the suggestion is generated:

      as the text content is received from the input unit; or

      after instructions to transmit the target content are received from the input unit.

5. The system of claim 1, wherein the memory comprises a classification model executable by the one or more processors to process the target content and evaluate compliance with a plurality of content compliance standards.

6. The system of claim 5, wherein the classification model further processes series of content related to the target content to evaluate compliance.

7. The system of claim 5, wherein the target content relates to a transaction and the classification model further processes transaction parameters to evaluate compliance.

8. The system of claim 5, wherein the memory comprises a suggestion content generation model executable by the one or more processors to generate the suggestion content based on an output of the classification model.

9. The system of claim 1, wherein the target content and the suggestion content comprise at least one of: text content or audio-visual content.

10. A computer-implemented method of content analysis executable by a system in communication with a remote device, the method comprising:

receiving a target content from a remote device;

processing the target content as it is received by the system to evaluate compliance with a plurality of content compliance standards;

on determining non-compliance of the target content, generating a suggestion content in real-time or near real-time; and

communicate the suggestion content to the remote device.

11. The method of claim 10, wherein the suggestion comprises compliant content generated based on the target content.

12. The method of claim 10, wherein the compliance of the target content is evaluated based on a series of content related to the target content, the series of content comprising content in a communication thread that the target content is a part of.

13. The method of claim 13, wherein the target content is text content; and

the target content is evaluated and the suggestion is generated:

as the text content is received from the remote device; or

after instructions to transmit the target content are received from the remote device.

14. The method of claim 10, a classification model is executed by the system to process the target content and evaluate compliance with a plurality of content compliance standards.

15. The system of claim 14, wherein the classification model further processes series of content related to the target content to evaluate compliance.

16. The system of claim 14, wherein the target content relates to a transaction and the classification model further processes transaction parameters to evaluate compliance.

17. The system of claim 14, wherein a suggestion content generation model is executed by the system to generate the suggestion content based on an output of the classification model.

18. The method of claim 10, wherein the target content and the suggestion content comprise at least one of: text content or audio-visual content.

19. A system comprising:

one or more processor (processor(s));

a memory comprising instructions that when executed by one or more processors cause the system to:

receive a target content from a remote device;

process the target content as it is received by the system to evaluate compliance with a plurality of content compliance standards;

on determining non-compliance of the target content, generate a suggestion content in real-time or near real-time; and

transmit the suggestion content to the remote device.

20. A computer-implemented method of content analysis executable by a system comprising an input device and an output device, the method comprising:

receive a target content from the input device;

process the target content to evaluate compliance with a plurality of content compliance standards;

on determining non-compliance of the target content, generate a suggestion content in real-time or near real-time; and

communicate the suggestion content to a user of the system using the output unit.