

Patent Cooperation Treaty Application**USA Receiving Office****Title: Pre-Publication Assessment of Digital Content****Inventor: Mustafa Behan****TECHNICAL FIELD**

[0001] The present disclosure relates to training and using machine learning assessment and decision models that analyze digital content from various origin points to identify fraudulent content prior to its publication.

BACKGROUND

[0002] Many online platforms use methods to identify, analyze and remove third-party (e.g., individuals, companies, AI) user generated fake and/or dis- (or mis-)informative digital content (fraudulent digital content). These approaches generally rely on post-publication review for authenticity decisions. Post-publication review does not properly consider the situation that once fraudulent digital content is released, if even for a short period of time, it has still been available for viewing and relied upon by at least some people. Further, post-publication review, due to the nature of the processes, often misses fraudulent digital content, thereby resulting in high percentages of fraudulent information remaining for viewing. Other methods used rely on semantic analysis of digital content that works pre-publication; however, these approaches detect fraudulent content from very simple manipulation schemes. These systems fail to detect content from more complex fraudulent content generation (e.g., conversational AI).

SUMMARY

[0003] Various embodiments of the disclosure relate to a machine learning method which may be implemented by a computing system of a data integrity systems operator. The method may comprise generating fraudulent digital content data markers by applying digital content to a model of known peer data points, the relationships between the known peer data points and the digital content to assess the potential for fraudulent digital content. The method may comprise sending the digital content to a machine learning analytical model to generate an authenticity assessment of certain digital content. The machine learning assessment model may have been trained by applying machine learning to analytical models and digital content to synthesize machine learning to understand large sets of digital content.

[0004] Various embodiments of the disclosure relate to a machine learning computing system comprising a processor and a memory comprising instructions executable by the processor. The instructions may comprise generating data markers for fraudulent digital content by processing the digital content in a model of known peer data points and identifying the relationships between them and the digital content. The machine learning platform may comprise sending the digital content to a machine learning assessment model to generate an assessment of fraudulent digital content. The machine learning assessment model may have been trained by applying machine learning to digital content and known peer digital content data points.

[0005] Various embodiments of the disclosure relate to a machine learning method which may be implemented by a computing system of a data integrity systems operator. The method may comprise generating digital content indicators by identifying

singularities to develop known peer data points and apply them to identify and assess fraudulent digital content. The method may comprise sending the digital content to a machine learning assessment model to generate an assessment of fraudulent digital content. The machine learning assessment model may have been trained by applying machine learning to network features and non-network features.

[0006] In various example embodiments, the network model may comprise a consumer review. The machine learning platform may define at least one singularity in the consumer review according to at least one of a distance between reviews and time lags between responses setting a fingerprint and identifying patterns between consumer review digital content and known peer digital content data points. The machine learning platform translates each consumer review into a metrical value to identify patterns and similarities within a set of consumer reviews. These patterns are translated into an assessment of similarities to understand common sources of content— which may then be used to identify fraud.

[0007] In various example embodiments, the network model may comprise a commentary. The machine learning platform translates each commentary into a metrical value to identify patterns and similarities within a set of commentaries. These patterns are translated into an assessment of similarities to understand common sources of information – which may then be used to identify fraud.

[0008] In various example embodiments, the model may comprise a retail online site that displays consumer reviews. The machine learning platform may translate each consumer review into a metrical value to identify patterns and similarities within a set of

consumer reviews. These patterns are translated into an assessment of similarities to understand common sources of information – which may then be used to identify fraud.

[0009] In various example embodiments, the network model may comprise a publicly available social commentary platform such as a social media platform, a news organization platform and the like. The machine learning platform may translate each social commentary into a metrical value to identify patterns and similarities within a set of social commentary. These patterns may then be translated into an assessment of similarities to understand common sources of information – which may then be used to identify fraud.

[0010] In various example embodiments, the model may comprise distance operators. Known data points may comprise distance points for determining distance and/or time between reviews and known peer data points. The machine learning platform may be configured to generate fraudulent digital content data markers by creating data points for combinations of distance operators and comparable known data points.

[0011] In various example embodiments, the model may comprise a network corresponding to linkages between known data points. The machine learning platform may translate each linkage into a metrical value to identify patterns and similarities within a set of linkages. These patterns may be translated into an assessment of similarities to understand common sources of information – which may then be used to identify fraud.

[0012] Various embodiments of the disclosure relate to a machine learning method implemented by a computing system of a data integrity system operator. The method may train one or more assessment models for assessing digital content to determine

its validity prior to its publication. The method may comprise generating a model of known peer data points and correlating relationships between digital content and the known peer data points. The network model may comprise at least one digital content unit (e.g., a consumer review, commentary posting, etc).

[0013] Each singularity in the digital content unit may be defined according to at least one of a distance between reviews and time lags between responses to follow-up plausibility questions setting a fingerprint and identifying patterns between generated digital content and known peer digital content data points. The method may comprise generating digital content data points by applying known peer data points to the model. The method may comprise applying machine learning to the known fraudulent peer digital content data points to train a machine learning assessment model configured to generate assessments of fraudulent digital content.

[0014] Various embodiments of the disclosure relate to a machine learning method implemented by a computing system for the pre-publication analysis and assessment of digital content, wherein pre-publication digital content is received by a computing system and is verified. A machine learning assessment model using known peer data points as training data is then engaged, the machine learning assessment model trained to generate pre-publication identification of fraudulent digital content. The machine learning assessment model, using the known data points as input to generate an assessment of digital content to determine whether it is fraudulent, the machine learning assessment model comprising a density-based clustering technique that is a function of a density operator. The machine learning assessment model defining each singularity in the network according to at least one of an amount, a frequency, or an incidence of data points between the received digital content and known peer data

points. The model comprises distance operators, and wherein generating the peer data records comprises generating, by the computing system, features for combinations of distance operators and distance vectors compatible with corresponding distance operators.

[0015] Various embodiments of the disclosure including a machine learning computing system for detection of fraudulent digital content, the computing system comprising a processor and a memory comprising instructions executable by the processor, the instructions comprising a machine learning platform configured to train a machine learning assessment model using known peer data points, in whole or in part, as training data, the machine learning assessment model trained to generate assessments of whether digital content is fraudulent. Execution by the machine learning assessment model using the known peer digital content as input to generate an assessment of whether digital content is fraudulent. The model screening pre-publication digital content at least once for fraudulent data points, wherein the machine learning platform defines each singularity in the digital data according to at least one of a frequency or an incidence of known peer data.

[0016] The model including a network corresponding to linkages between known peer data records, including data points, wherein the machine learning platform defines each singularity in the network as indicating a linkage between corresponding tuples of known peer digital content data points and pre-publication digital content, each singularity weighted according to a characteristic of the corresponding linkage.

[0017] These and other features are explained in more detail in and will become apparent from the following detailed description and the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0018] **FIG. 1** is a flow diagram of a general overview of the system.
- [0019] **FIG. 2** is a block diagram of a data integrity system operator computing system that implements a machine learning platform and that communicates with third-party platforms.
- [0020] **FIG. 3** is a process flow diagram for a machine learning approach to detecting fraudulent digital content.
- [0021] **FIG. 4** is a decision tree diagram illustrating the steps of digital content analysis.

DETAILED DESCRIPTION

- [0022] In the various embodiments presented herein, machine learning approaches, which may involve supervised and/or unsupervised learning, may be used for training and implementing assessment machine learning models for detecting fraudulent information or misinformation a/k/a disinformation prior to its publication. Throughout, the term fraudulent digital content includes, without limitation, fake or untruthful information and mis- or dis-information. Fraudulent digital content may, without limitation, arise in the settings of reviews, such as found on shopping, travel, hospitality, and restaurant and restaurant related sites, and commentary, such as publicly accessible news, social media, and politically oriented sites. Throughout digital

content includes, without limitation, user and/or machine generated reviews and commentary created for the purpose of publishing on the internet to websites or platforms. Fraudulent digital content may originate from, without limitation, individuals, companies, paid content creators, and/or machine learning/artificial intelligence sources.

[0023] A representation of digital content and its relationships to a known peer group of digital content data points can be determined and used to translate identified digital content data points into bench-marking peer data points. Machine learning models may be applied to analyze connections and understand relations between different data points and thereby more effectively and efficiently discover fraudulent digital content prior to its publication.

[0024] A machine learning platform trains and uses models to exploit the fact that certain digital content may contain data points that identify them as fraudulent. Each singularity of the digital content may be defined as containing a linkage between known fraudulent data points which, in turn, may be weighted according to one or more characteristics of the known peer group data points. A singularity indicating a linkage between corresponding tuples of known peer digital content data points and pre-publication digital content, wherein the computing system defines each singularity as denoting a linkage between corresponding tuples, pairs are sets of two, n-tuples are set of n elements of digital content, and each singularity weighted according to a characteristic of the corresponding digital content. The method may comprise generating fraudulent data points by applying known peer data points to the model. The method may comprise applying machine learning to the known peer group data points to train a machine learning assessment model configured to generate

assessments of fraudulent digital content, and which may be described by distance operators.

[0025] In various example embodiments, the method may comprise comparing digital content data points corresponding to known peer group data points to generate an assessment of fraudulent digital content. This process is undertaken after the time of digital content creation but before a potential publication.

[0026] The model may comprise curved-space distance operators. Unlike orthogonal matrices which generate results based only on angled vectors of 90 degrees thus creating a square matrix used to represent a finite graph, the herein developed and applied distance operators enable a broader scope of activity by enabling a curved inclusion area, thereby offering the opportunity for additional data point recognition. The use of these distance operators further enhances the ability to interpret digital content data points using multiple peer data points and known fraudulent data points. The known fraudulent data points may comprise codifying from combinations of distance operators and corresponding known peer group data points to identify similarities in certain digital content.

[0027] The machine learning platform may employ a variety of machine learning techniques in classification and pattern recognition of digital data points to determine “organic” (as opposed to “fabricated”) patterns in data and to identify unusual outlier data points.

[0028] Fraudulent digital content is often not discoverable through analyzing the individual digital content in isolation, but through data point relationships and linkages with peer group data points. Whether the peer group data points are fraudulent or not

is immaterial; the pattern/distance between data points provides the necessary information to make an assessment. Certain activities may disguise fraudulent digital content unless the activities are placed in the context of known peer group data points with some identifiable association which may, in isolation, seem normal. The disclosed approach thus enables the discovery and prevention of fraudulent digital content that could not otherwise be discovered prior to publication through conventional means. Further, the disclosed approach reveals relevant connections within digital content not feasibly discoverable by human users. Training an assessment model using a combination of network-based and other features as disclosed herein places the digital content in context rather than being considered in isolation, thereby yielding accurate assessments.

[0029] Example embodiments of the machine learning model described herein improve computer-related technology by performing functions that cannot be carried out by conventional computing systems. Furthermore, the embodiments described herein cannot be carried out by humans. The machine learning model may proactively compare digital content corresponding to known peer group data points and known fraudulent data points to identify fraudulent digital content that would otherwise go undetected. Conventional systems may include databases and definitions that are static and cannot be configured to acquire, store, and update the multitude of information required from identifying peer group data points and known fraudulent data points.

[0030] Referring to FIG. 1, a flow diagram illustrating the basic movement of digital content through the inventive system **100** is shown, according to potential embodiments.

[0031] The system **100** includes a platform computing system **102**, here a data integrity operator (DIO). A DIO may be defined as a centralized server that houses peer group data points in a Database **116**. The platform computing system **102** may be that of an online provider of goods or services or a platform that accepts user created digital content. For example, many online platforms accept user-created digital content in the form of reviews of goods or services, such as online shopping platforms, travel related platforms, restaurants and restaurant related platforms, and the like. Other platforms accept user created commentary, such as news organizations, social media platforms, political platforms, and the like. While the DIO may comprise an online platform such as a retailer or news organization, various embodiments contemplate centralized servers that house peer group data points, such as a server farm. The DIO **102** further comprises memory hosting a database **116** for housing the data points and a DIO Management Module **118** for operating the system. Database **116** and Management Module **118** communicate with DIO **102** at communication points **102a** and **102b**, respectively.

[0032] The component parts of the system **100** may be integrated or operatively coupled to one another directly or over a network that permits the exchange of data and the like. System **100** may include one or more processors, memories, network interfaces, and user interfaces. The memory may also store data points in database **116**. Interface **106** allows system **100** to communicate **104** with DIO **102**, when operatively coupled, by sending and receiving transmissions via one or more communications protocols.

[0033] Once a platform computing system **102** communicates **104** at Interface **106**, the digital content submission and analysis process is initiated by the computing system.

Review Collection Module **110** organizes the way users can submit reviews and plausibility check **120** is initiated wherein information is collected, such as, but not limited to, time of submission, IP address, email address, topic of digital content, keywords, etc. Simultaneously or later, Review Collection Module **110** and then may send **112** digital content to Review Collection Widget **114**. Review Collection Widget **114** comprises a link or a link on the Website of the platform **102** to signal to users that they can place/submit their digital content here. The plausibility check **120** is an order check of the digital content itself, without comparison to any other content or data point. The plausibility check may happen through an email that asks plausibility questions regarding the digital content submission. The computing system may then check between what is written in the review and the answers of the plausibility question answers. Plausibility is determined by analyzing basic features of the digital content such as consistency in subject, email address, etc. A digital content submission may fall out or move on further in the analysis process. In either event, verification check **122** is initiated. For reviews that are disqualified in the plausibility check, data points are collected and stored in data base **116** and made accessible to the machine learning system for training and that can be used to better understand better how fraudulent content looks. Subsequent content items may then be compared to disqualified items and their characteristics. Verification checks are well known in the industry. Digital content is verified by (1) identifying the IP address of the digital content and, following, (2) sending a verification to the sender asking about the digital content. Verification is generally done via email but may be done through any system, such as text, DM, or telephone.

[0034] The present invention places metrics onto data points that until now have normally not had metrics. Digital content features such as word structure, sentence structure, proximity checks, characterization similarities, among other digital content features along with time lags (time patterns) are recorded and analyzed in combination with and against one another to locate patterns in digital content that identify fraudulent digital content. For non-limiting example, if multiple digital content units are received, all having the same or similar time lags sentence structure, suspicion is raised about the validity of the digital content. Even if time lags are varied, patterns emerge that signal potential fraudulent digital content which the machine learning computing system can detect and store for future digital content analysis. This true for all features in isolation or in combination. It is for this reason that the present invention works particularly well in identifying fraudulent digital content generated by artificial intelligence platforms.

[0035] Once the first plausibility check is completed, a first order analysis **124** is performed by the model wherein data points for each digital content are codified and then, using the different known peer group data points compared to other known peer group data points for plausibility of the digital content through measurement of the distance between other comparable digital content – and consistency within the verification. A probability of valid digital content is determined by measuring the distance of data points against known peer group data points. Through multiple iterations, at least two but that may number up to 70 or more, further determination can be made.

[0036] Further analysis is initiated by the model in a second order analysis **126** wherein measurement may identify patterns between groups of digital content. The second

order analysis may look at any forms of patterns, be it in time, text, etc. For example, a machine that writes content would create patterns that could be identified. Such analysis is not easy to identify in one single piece of content, but by looking at patterns in the peer group. Again, the second order analysis is performed through multiple iterations numbering up to 70 or more.

[0037] The outcome of the analysis is then weighted against known peer group data points to make an assessment **128** of the validity of the digital content. The machine learning model of the present invention is premised on the hypothesis that:

- (i) fraudulent digital content is fabricated,
- (ii) the source of fraudulent digital content is often originally the same,
- (iii) a common source leaves traces in characteristics in the digital content; and,
- (iv) assessment of the probability of fraud can be accomplished by detecting traces for a common source of the reviews in question.

[0038] With the above as a basis, the machine learning model of the present invention includes the following protocols:

1. Digital Content collection process: Process to collect digital content in a way that (i) ensures a high likelihood of authenticity at the collection stage and (ii) generation of an array of characteristics for each digital content;

2. Initial Assessment of Authenticity: Identification of red flags or white flags connected with the digital content, including:

- (I) history of reviewed vendor, product or service and
- (ii) history of digital content creator;

3. Measuring Characteristics of the digital content in Question: Isolate and classify single characteristics and establish a metric operator for the each of the single characteristics;
4. Classify the Intrinsic Authenticity of each new Single Digital Content: Identify intrinsic characteristics of each digital content to obtain a first estimate of authenticity and decide for specific modules of analysis to be used.
5. Measuring the Distance between the Digital Content in Question and other comparable Digital Content: Calculate and normalize the distance between single digital content within a given population to identify patterns between digital content data points;
6. Assess the Authenticity of the new incoming Single Digital Content: Combine the intrinsic plausibility of the single digital content with the distance assessment of other digital content in order to make an assessment of authenticity of digital content data points;
7. Deploy Machine Learning Techniques for Optimizing Accuracy: Increase the accuracy of the analysis of digital content by fortifying digital content trust points and identifying single distance patterns between digital content as suspicious.

[0039] Turning to **Fig. 2**, is a depiction of a system **200** flow example of a DIO working with a third-party site that displays user content via a UI website. Third party **202** can be any site or platform that displays user-generated digital content. Common sites include retailers, hospitality providers, travel sites, aggregation sites, news, social media, political sites, and others. These sites may provide specific locations for user created digital content (e.g., retail shopping sites enabling reviews of products) or may

be designed to generally accept comment (e.g., social media platforms) **204**. Third party **202** contacts **206** DIO website through its landing page **208**. Third party **202** interfaces DIO through a UI wherein it signs up with DIO and submits digital content criteria, among other things **210a**. Once a digital content submission is verified and activated **210B**, third party **202** sends DIO digital content for analysis **214** using the machine learning model **216**. Mapping of third-party digital content **216a** is performed by the model using DIO's data point data base as implemented by a computing system employing the machine learning model, including plausibility analysis **216b**, first order analysis **216d** and second order analysis **216d**, as discussed above, is performed by the computing system (see discussion above, **Fig. 1**). A determination is made as to whether the digital content is fraudulent **216d**. The determination is then transmitted **218** to the third party **202** as feedback **220**, **222** through third party's website and then to DIO dedicated dashboard **222b**. Third party retrieves the feedback from its dashboard and makes an internal decision or use the proposed decision developed by the computing system whether to publish the digital content.

[0040] Turning to **Fig. 3**, a process flow diagram for a machine learning approach to detecting fraudulent digital content is depicted. Machine learning system **300** obtains content to be checked **302** and performs a plausibility check **304**, as discussed above in reference to **Fig. 1**, of the digital content. The digital content is passed **306** for digital content synthesis **308** wherein digital content is codified into defined dimensions for further analysis. Machine learning system **300** then transfers **310** synthesized digital content for dissection and mapping **312**. Digital content is dissected **312a** into single review vectors, i.e., data points. Thereafter, the single review vectors are mapped against DIO numerical review model **312b**, i.e., known peer group data points for

analysis **312c** of patterns between single digital content vectors to benchmark then with known peer group content. Once pattern and bench-marking analysis is complete, machine learning system **300** then analyzes the result **316** and makes an assessment as to the authenticity of the digital content **316a** and a determination of whether the digital content should be published **316b**.

[0041] Thereafter, machine learning system **300** provides feedback of the assessment to the peer group database to train the model **316c** while also providing feedback to the requesting website **316d**.

[0042] Turning to **FIG.4**, an example decision tree of the machine learning model is illustrated. At **402**, a user submits digital content for potential publication. Certain information is collected by the model. This may be an email address, IP address, text (which may include keywords, titles, product or service identifier, etc.). Initial analysis is performed at **404** to determine whether the submission comes from a verified user. If so **408**, digital content is published **410** and content characteristics are stored within the database **116** (**FIG.1**) for machine learning model learning purposes and later analysis integration **412**. If digital content is not from a verified user **413**, it will undergo further analysis **414** as described above. Plausibility questions will be sent to the user **416**. If plausibility questions are not answered **418**, digital content will be blocked. Since no answers were provided to plausibility questions, content characterizations will not be saved. If plausibility questions are answered **422**, then an analysis is done by the computing system to determine whether the provided answers are plausible. If not **424**, digital content will be blocked, and content characteristics will be stored within the database **116** (**FIG.1**) for machine learning model learning purposes and later analysis integration **426**. If digital content is plausible **428**, it undergoes further analysis **430**

where content characteristics are made into data points and checked against known peer group data points, i.e., first order analysis.

[0043] If digital content characteristics (data points) are similar to peer group authentic review known data points **444** or to known fraudulent digital content data points **446**, digital content will be blocked, and content characteristics will be stored within the database **116 (FIG.1)** for machine learning model learning purposes and later analysis integration **447**. In the case that digital content data points are different from authentic known peer group data points and known fraudulent data points **432**, the digital content data points are further checked for content characteristics within known peer groups **434**, i.e., second order analysis. If digital content characteristics are not part of a pattern of authentic or fraudulent digital content peer groups **436**, it will be published **438** and content characteristics will be stored within the database **116 (FIG.1)** for machine learning model learning purposes and later analysis integration. In the cases where digital content data point characteristics are close to a pattern of known authentic data points **440** or known fraudulent digital content data points **442**, digital content will be blocked, and content characteristics will be stored within the database **116 (FIG.1)** for machine learning model learning purposes and later analysis integration **446**.

[0044] At machine learning process steps **432, 434**, multiple iterations are undertaken. Typically, there are between 70 and 200 data points per peer digital content. Each new submission adds to that count and each new submission is reviewed against all known group peer reviews. Thus, each new digital content submission may be analyzed at each of these two steps between 70 and 200 times multiplied by the number of reviews

in the peer group. Though the range of 70 to 200 is typical, it is not meant to be limiting.

[0045] The embodiments described herein have been described with reference to drawings. The drawings illustrate certain details of specific embodiments that provide the systems, methods and programs described herein. However, describing the embodiments with drawings should not be construed as imposing on the disclosure any limitations that may be present in the drawings.

[0046] The foregoing description of embodiments has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the disclosure to the precise form disclosed, and modifications and variations are possible in light of the above teachings or may be acquired from this disclosure. The embodiments were chosen and described in order to explain the principals of the disclosure and its practical application to enable one skilled in the art to utilize the various embodiments and with various modifications.

[0047] Example computing systems and devices may include one or more processing units each with one or more processors, one or more memory units each with one or more memory devices, and one or more system buses that couple various components including memory units to processing units. Each memory device may include non-transient volatile storage media, non-volatile storage media, non-transitory storage media (e.g., one or more volatile and/or non-volatile memories), etc. In this regard, machine-executable instructions comprise, for example, instructions and data which cause a general-purpose computer, special purpose computer, or special purpose processing machines to perform a certain function or group of functions. Each

respective memory device may be operable to maintain or otherwise store information relating to the operations performed by one or more associated modules, units, and/or engines, including processor instructions and related data (e.g., database components, object code components, script components, etc.), in accordance with the example embodiments described herein.

[0048] It should be noted that although the drawings herein may show a specific order and composition of method steps, it is understood that the order of these steps may differ from what is depicted. For example, two or more steps may be performed concurrently or with partial concurrence. Also, some method steps that are performed as discrete steps may be combined, steps being performed as a combined step may be separated into discrete steps, the sequence of certain processes may be reversed or otherwise varied, and the nature or number of discrete processes may be altered or varied. The order or sequence of any element or apparatus may be varied or substituted according to alternative embodiments. Accordingly, all such modifications are intended to be included within the scope of the present disclosure as defined in the appended claims. Such variations will depend on the machine-readable media and hardware systems chosen and on designer choice. It is understood that all such variations are within the scope of the disclosure. Likewise, software and web implementations of the present disclosure may be accomplished with standard programming techniques with rule-based logic and other logic to accomplish the various database searching steps, correlation steps, comparison steps and decision steps.

[0049] The foregoing description of embodiments has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the disclosure

to the precise form disclosed, and modifications and variations are possible in light of the above teachings or may be acquired from this disclosure. The embodiments were chosen and described in order to explain the principals of the disclosure and its practical application to enable one skilled in the art to utilize the various embodiments and with various modifications as are suited to the particular use contemplated. Other substitutions, modifications, changes and omissions may be made in the design, operating conditions and arrangement of the embodiments without departing from the scope of the present disclosure as expressed in the appended claims.

What is claimed:

1. A machine learning method implemented by a computing system for pre-publication assessment for fraudulent digital content, the method comprising:

receiving, by the computing system, pre-publication digital content;

verifying the digital content;

training, by the computing system, a machine learning assessment model using known peer group data points as training data, the machine learning assessment model trained to conduct generate pre-publication identification of fraudulent digital content;

generating, by the computing system, based on a plurality of data records accessed via an electronic database, a model including a set of patterns and relationships there between of learned data points known to generate fraudulent digital content;

retrieving, by the computing system, from the electronic database, a plurality of the known peer group data points;

generating, by the computing system, an assessment analysis of fraudulent digital content by screening the digital content against the peer group data records, the screening multiplied according to a plurality of iteration steps;

executing, by the computing system, the machine learning assessment model using the known data points as input to generate an assessment of a fraudulent digital content, the machine learning model comprising a density-based clustering technique that is a function of a density parameter; and

generating, by the computing system, a marker to the identification of fraudulent digital content.

2. The machine learning method of claim 1, wherein the digital content is selected from the group comprising consumer reviews, political comment, and healthcare information.

3. The machine learning method of claim 2, wherein the computing system defines each singularity in the network according to at least one of a time pattern, an amount, a frequency, or an incidence of data points between the received digital content and peer group data.

4. The machine learning method of claim 3, wherein generating the model comprises detecting, by the computing system, known fraudulent data points in the peer group data and screening them against the digital content for similarities in data points.

5. The machine learning method of claim 4, wherein digital content that includes potential fraudulent data points is marked for removal.

6. The machine learning method of claim 5, wherein newly identified fraudulent data points are included in the peer group data records.

7. The machine learning method of claim 6, wherein generating the model comprises detecting, by the computing system, common relationships to assess digital content sources.

8. The machine learning method of claim 1, wherein the model includes similarities in content between digital content further comprising time lags between receiving digital content and verifying the digital content.

9. The machine learning method of claim 8, wherein the computing system defines each singularity as denoting a linkage between corresponding tuples, pairs are sets of two, n-tuples are set of n elements of digital content, each singularity weighted according to a characteristic of the corresponding digital content.

10. The machine learning method of claim 9, wherein generating the model comprises detecting, via the computing system, linkages using data accessed via stored peer group data records and internet based platforms that publish digital content to identify similarities between digital content.

11. The machine learning method of claim 1, wherein the machine learning platform defines each singularity as content vectors comprising fraudulent content.

12. The machine learning method of claim 11, wherein the model comprises distance operators, and wherein generating the peer group data records comprises generating, by the computing system, features for combinations of distance operators and data points compatible with corresponding distance operators.

13. A machine learning computing system for detection of fraudulent digital content, the computing system comprising a processor and a memory comprising instructions executable by the processor, the instructions comprising a machine learning platform configured to:

train, by the computing system, a machine learning assessment model using peer group digital content, in whole or in part, as training data, the machine learning assessment model trained to generate assessment of fraudulent digital content and issue markers identifying potential fraudulent digital content,

generate, based on a plurality of data records, a model including known fraudulent digital content and relationships there between;

execute the machine learning assessment model using the peer group digital content as input to generate an assessment of fraudulent digital data, the machine learning assessment model comprising a density-based clustering technique that is a function of a density parameter; and,

generate a marker in response to the assessment of the fraudulent digital content.

14. The machine learning system of claim 13, wherein the model comprises pre-publication digital content screened at least once for fraudulent data points, and wherein the machine learning platform defines each singularity in the digital data according to at least one of a frequency or an incidence of known fraudulent peer group data.

15. The machine learning system of claim 13, wherein the digital content is selected from the group comprising consumer reviews, political comment, and healthcare information.

16. 16. The machine learning system of claim 13, wherein the model comprises linkages between peer group data point records and digital content data points, and wherein the machine learning platform defines each singularity as denoting a linkage between corresponding pairs of peer group digital content and pre-publication digital content, each singularity weighted according to a characteristic of the corresponding linkage.

17. A machine learning computing system for detection of fraudulent digital content, the computing system comprising a processor and a memory comprising instructions executable by the processor, the instructions comprising a machine learning platform configured to:

train, by the computing system, a machine learning assessment model using peer group digital content, in whole or in part, as training data, the machine learning assessment model trained to generate assessment of fraudulent digital content and issue markers identifying potential fraudulent digital content,

generate, based on a plurality of data records, a model including known fraudulent digital content and relationships there between;

execute the machine learning assessment model using the peer group digital content as input to generate an assessment of fraudulent digital data, the machine learning assessment model comprising a density-based clustering technique that is a function of a density parameter employing distance operators; and,

generate a marker in response to the assessment of the fraudulent digital content.

18. The machine learning computing system of claim 17, wherein the model further includes authentic digital content.

19. The machine learning system of claim 17, wherein the machine learning assessment model executes a first order analysis wherein data points for each digital content are codified and, using known peer group data points, compared to other known peer group data points for plausibility of the digital content through multiple iterations, the machine learning assessment model then executes a second order analysis wherein, through multiple iterations, patterns between groups of digital content are identified through detection of patterns

between other comparable digital content, a probability of valid digital content being determined by measuring the distance of data points against known peer group digital content.

20. The machine learning assessment model of claim 19, wherein each digital content is codified into data points and analyzed against all known group peer reviews, each digital content submission analyzed against known data points multiplied by the number of reviews in the peer group.