

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

CROWDSTRIKE, INC.,
Petitioner,

v.

SKYSONG INNOVATIONS, LLC,
Patent Owner.

PTAB Case No. IPR2025-01398
U. S. Patent No. 11,275,900

PATENT OWNER'S PRELIMINARY RESPONSE

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	RELATED PROCEEDINGS	1
III.	SKYSONG BACKGROUND	1
IV.	THE ASSERTED PRIOR ART	2
A.	<i>Deliu</i> (EX1004)	3
B.	<i>Pendar</i> (EX1005)	5
C.	<i>Frank</i> (EX1007)	6
D.	<i>Ramírez-Corona</i> (EX1008)	7
E.	<i>Chawla</i> (EX1009).....	7
F.	<i>Xiao</i> (EX1010).....	8
G.	<i>Ukrainczyk</i> (EX1011)	8
H.	<i>Nunes</i> (EX1012)	9
V.	THE '900 PATENT	10
A.	Technical Background.....	10
B.	Prosecution History	11
C.	Priority Date	12
VI.	LEVEL OF ORDINARY SKILL IN THE ART	13
VII.	PATENT OWNER'S CLAIM CONSTRUCTIONS	13
VIII.	PETITIONER HAS FAILED TO DEMONSTRATE A REASONABLE LIKELIHOOD OF PREVAILING AS TO ANY CHALLENGED CLAIM.	14

A.	GROUND 1: Obviousness in Light of <i>Deliu</i> and <i>Pendar</i>	14
1.	Claim 1.....	14
2.	Claim 2.....	27
3.	Claim 4.....	28
4.	Claim 5.....	29
5.	Claim 10.....	29
6.	Claim 12.....	29
7.	Claim 13.....	30
8.	Motivation To Combine.....	32
B.	Ground 2: Obviousness in Light of <i>Deliu</i> , <i>Pendar</i> , and <i>Frank</i>	35
1.	Claim 3.....	35
2.	Motivation To Combine.....	37
C.	Ground 3: Obviousness in Light of <i>Deliu</i> , <i>Pendar</i> , and <i>Ramírez-Corona</i>	37
D.	Ground 4: Obviousness in Light of <i>Deliu</i> , <i>Pendar</i> , and <i>Chawla</i>	38
1.	Claim 6.....	38
2.	Claim 14.....	38
E.	Ground 5: Obviousness in Light of <i>Deliu</i> , <i>Pendar</i> , and <i>Xiao</i>	39
1.	Claim 7.....	39
2.	Motivation To Combine.....	40
F.	Ground 6: Obviousness in Light of <i>Deliu</i> , <i>Pendar</i> , and <i>Ukrainczyk</i>	41
1.	Claim 8.....	41

2.	Claim 9.....	41
G.	Ground 7: Obviousness in Light of <i>Deliu</i> , <i>Pendar</i> , and <i>Nunes</i>	42
1.	Claim 11.....	42
IX.	SECONDARY CONSIDERATIONS OF NON-OBVIOUSNESS	42
A.	Commercial Success.....	42
B.	Long-Felt Need and Government Support	44
C.	Failure Of Others.....	45
D.	Unexpected Results	45
E.	Skepticism of Experts.....	45
F.	Praise By Others	46
G.	Copying By Competitors.....	46
X.	CONCLUSION.....	46

TABLE OF AUTHORITIES

CASES

<i>Graham v. John Deere Co. of Kan. City</i> , 383 U.S. 1 (1966).....	64
<i>Phillips v. AWH Corp.</i> , 415 F.3d 1303 (Fed. Cir. 2005) (<i>en banc</i>)	13

TABLE OF ABBREVIATIONS

Patent Owner	Skysong Innovations, LLC (“Skysong”)
Petitioner	CrowdStrike, Inc. (“CrowdStrike”)
PTAB	Patent Trial and Appeals Board
’900 patent	U.S. Patent No. 11,275,900 (EX1001)
<i>Pendar</i>	U.S. Patent No. 9,330,167 to Pendar (EX1005)
<i>Frank</i>	Eibe Frank, Pruning Decision Trees and Lists, University of Waikato, submitted January 2000 (EX1007)
<i>Ramírez-Corona</i>	Mallinali Ramirez-Corona et al., <i>Multi-label Classification for Tree and Directed Acyclic Graphs Hierarchies</i> , Probabilistic Graphical Models, PGM 2014, (2014) (EX1008)
<i>Chawla</i>	Nitesh V. Chawla et al., <i>SMOTE: Synthetic Minority Oversampling Technique</i> , Journal of Artificial Intelligence Research 16 (2002) 321-357 (EX1009)
<i>Xiao</i>	Tong Xiao et al., <i>KNN and Re-ranking Models for English Patent Mining at NTCIR-7</i> , Proceedings of NTCIR-7 Workshop Meeting, (2008) (EX1010)
<i>Ukrainczyk</i>	U.S. Patent No. 7,028,250 to Ukrainczyk et al. (EX1011)
<i>Nunes</i>	Eric Nunes et al., <i>Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence</i> , Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics (EX1012)

LIST OF EXHIBITS

Exhibit No.	Descriptions
2001	Skysong's First Amended Complaint against CrowdStrike
2002	Declaration of Dr. Paulo Shakarian regarding discretionary denial with Appendix
2003	Civil Docket For Case # 7:25-cv-00040-DC-DTG
2004	<i>Skysong v. CrowdStrike</i> , Dkt. No. 55, Scheduling Order
2005	Table of CrowdStrike IPRs and Parallel District Litigation Schedules
2006	Civil Docket For Case # 2:25-cv-00098-JRG
2007	Revision to Rules of Practice Before the Patent Trial and Appeal Board
2008	Executive Order No. 14,028
2009	Redacted Declaration of Dr. Paulo Shakarian regarding discretionary denial with Appendix
2010	Default Protective Order
2011	Declaration of Dr. Nenad Medvidovic

LIST OF CHALLENGED CLAIMS

Limitation	Claim Language
Claim 1	
1	A computer-implemented system for improving classification of criminal activities arising from cyber environments, comprising:
1.a	a processor configured to:
1.a.1	access a ground truth dataset generated from deep web forum information, the ground truth dataset defining a predetermined tag hierarchy including predetermined tag labels corresponding [to] topics of the deep web forum information;
1.a.2	access data related to a discussion topic for classification from a deep web forum;
1.a.3	extract a set of features from the data by assigning word vectors or paragraph vectors to the data; and
1.a.4	apply a machine classifier to the set of features to generate a prediction list of tags for classifying the discussion topic, wherein the prediction list includes a prediction probability value for each tag of the plurality of tags; and
1.a.5	add all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value.
Claim 2	
2	The computer-implemented system of claim 1, wherein a predetermined tag hierarchy is divided into parent tags and child tags, wherein a topic represented by a child tag is representative of a subset of a topic represented by a parent tag, wherein each child tag corresponds to a particular parent tag, and each tag is assigned a corresponding probability value.
Claim 3¹	
3	The computer-implemented system of claim 1, wherein the processor is configured to remove all child tags of the plurality of tags corresponding to a parent tag from the prediction list of tags where a probability value associated with the ~~child~~ [parent] tag is below a second predetermined threshold value.

¹ Claim 3 has been edited to correct the indicated typographical error.

Claim 4²	
4	The computer-implemented system of claim 1, wherein the processor is configured to add all parent tags of the plurality of tags corresponding to a child tag if a probability value associated with the parent [child] tag is above the first predetermined threshold value.
Claim 5	
5	The computer-implemented system of claim 1, wherein the set of features are extracted using textual feature extraction to identify tags associated with topic titles of the data.
Claim 6	
6	The computer-implemented system of claim 1, wherein a set of synthetic data samples are created for addition to the ground truth dataset to address class imbalance by creating a set of synthetic sample points along a line defined between a first minority feature vector and a second feature vector and adding the set of synthetic sample points to the ground truth dataset.
Claim 7	
7	The computer-implemented system of claim 1, wherein the ground truth dataset is supplemented by, for every sample of the ground truth dataset in a minority class, adding top similar documents from the deep web forum information, as identified using an elastic search similarity score, and adding them to the ground truth dataset.
Claim 8	
8	The computer-implemented system of claim 1, wherein the data is parsed and segmented into tokens, and predetermined characters and predetermined common words are removed from the data.
Claim 9	
9	The computer-implemented system of claim 1, wherein words within the data are reduced to their root form using stemming and lemmatization procedures.
Claim 10	
10	The computer-implemented system of claim 1, wherein the set of features are extracted from the data using a neural network, wherein the neural network is configured to learn a continuous distributed vector representation for documents and words within the data.

² Claim 4 has been edited to correct the indicated typographical error.

Claim 11	
11	The computer-implemented system of claim 1, further comprising: a crawler in operable communication with the processor and operable to traverse a plurality of deep web websites to retrieve Internet documents forming at least a portion of the deep web forum information.
Claim 12	
12	A method, comprising:
12.a	configuring a processor for executing operations including:
12.a.1	accessing data associated with a deep web forum, the data defining a topic for classification;
12.a.2	extracting a set of features from the data as inputs for a machine classifier; and
12.a.3	apply a machine classifier to the set of features to generate a prediction list of tags for classifying the topic, wherein the prediction list includes a prediction probability value for each tag of the plurality of tags; and
15.a.4	adding all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value.
Claim 13	
13	The method of claim 12, further comprising: outputting an update to the prediction list of tags, a total number of the tags of the prediction list of tags being different from the update to the prediction list of tags to maintain a predetermined tag hierarchy.
Claim 14	
14	The method of claim 12, further comprising generating a predetermined tag hierarchy based on a ground truth dataset, and leveraging semi-supervised learning to address class imbalance and increase accuracy of the machine classifier.

I. INTRODUCTION

Patent Owner Skysong respectfully submits this Preliminary Response to the Petition for *Inter Partes* Review dated Aug 25, 2025 (“Petition”) of U.S. Patent No. 11,275,900 (EX1001, “’900 patent”). This response is timely filed. *See* Paper 3.

II. RELATED PROCEEDINGS

The ’900 patent is the subject of the following related litigation:

- *Skysong Innovations, LLC v. CrowdStrike, Inc. and CrowdStrike Holdings, Inc.*, Case No. 7:25-cv-00040, filed January 31, 2025 in the United States District Court for the Western District of Texas; and
- *Skysong Innovations, LLC v. Fortinet, Inc.*, Case No. 2:25-cv-00098, filed January 31, 2025, in the United States District Court for the Eastern District of Texas.

III. SKYSONG BACKGROUND

Skysong is a nonprofit organization, and the exclusive intellectual property management and technology transfer organization for Arizona State University (“ASU”). Skysong works closely with ASU and its technology industry partners to translate ASU’s innovations into commercial applications, fostering economic development and public benefit. *See* EX2001 (Skysong’s First Amended Complaint), ¶¶3–4. Skysong significantly contributes to the mission statement of

the University and has become a core component of Arizona's innovation ecosystem and a driver of high-tech industry in the state.

The '900 patent stems from research conducted by Dr. Paulo Shakarian and his team. *See* EX2002, ¶¶5–6, 11. It discloses methods for classifying data associated with the deep web or dark net and concerning criminal activities arising from cyber environments. *See* EX1001, abstract.

IV. THE ASSERTED PRIOR ART

Dr. Nenad Medvidovic, Department Chair of the Computer Science Department of the University of Southern California and leading artificial intelligence and cybersecurity authority, has been retained by Skysong as an independent expert to assist in the technical analysis of the '900 patent and the asserted prior art. He has provided a declaration (EX2011, "*Declaration Of Nenad Medvidovic, Ph. D., In Support Of Patent Owner's Preliminary Response*") setting forth his opinions, including a synopsis of the relevant aspects of the asserted prior art. His synopsis of the asserted prior art is relevant and reproduced here for reference:

A. *Deliu* (EX1004)³

Deliu refers to Isuf Deliu et al., *Extracting Cyber Threat Intelligence From Hacker Forums*, NTNU Norwegian University of Science and Technology, doctoral thesis submitted June 2017.

Deliu is a comparative study of enhancements to traditional security controls. Specifically, *Deliu* explores the capabilities of a combination of supervised and unsupervised machine learning to locate relevant threat intelligence from hacker forums. *Deliu* posed two primary research question: “1. How can different Machine Learning methods be used to classify the contents of hacker forums?” and “2. What are the main topics related to cyber security on such forums, and how does filtering through classification affect the discovered topics?” *Deliu* at 19.

To address these questions, *Deliu* compared the classification performance of [then] recent deep learning algorithms with more traditional machine learning methods in a both a binary- and multi-class dataset constructed from data collected from a real hacker forum. *Deliu* at 19. According to *Deliu*, “this comparison is a novel contribution to the cyber security community.” *Deliu* focuses on the efficiency of the “how” to classify, as opposed to the final structure of the classification itself, i.e., whether it is hierarchical or not. *Deliu* concludes that “the traditional SVM classifier with character n-gram features perform at least as good

³ EX2011 at ¶¶83–91 (*Deliu*).

as the ConvNN.” *Id.* at 103. *Deliu* thus only compared the performance of various machine learning approaches with each other.

Deliu identifies no motivation to increase the complexity of the classification scheme (e.g., to a hierarchical scheme). Instead, *Deliu* explicitly teaches that SVM’s are preferred. EX1016 at 8 (“Given the computational complexity of CNN architectures, our results indicate that SVMs are superior for the purposes of practical, real-time CTI applications.”)

“*Traditional SVM*” is a binary classification model. *See Deliu* at 55 (*Nunes*’ uses “binary classification to categorize the hacker forums posts and marketplace products into two classes: (a) relevant and (b) irrelevant to cyber security.”) and 104 (“*Nunes et al.* used *traditional SVM* classifier with n-grams as features in a similar dataset.”) (emphasis added). Thus, *Deliu* was motivated to use the same binary classification model to be able to compare equal paradigms.

Deliu’s second goal aims at determining what the “main topics related to cyber security” are on such forums. This explicitly focus on “main topics” and would not benefit from the additional complexity of adding “sub-topics” in a topic hierarchy.

Deliu further confirms that multi-class data sets increase the complexity of the problem. *See, e.g., Deliu* at 106 (“Manual labeling might be resource demanding especially in the case of the multi-class dataset”); *Id.* at 63 (“We must emphasize that this kind of [multi-class] labeling is much more demanding.”); *Id.* at 74 (“the

construction of the [multi-class] dataset is much more demanding.”). Notably, imposing a hierarchical classification structure is non-trivial and further increases the complexity of the problem identified by *Deliu*.

Deliu thus teaches away from using a hierarchical classification scheme because it seeks to (1) use the same paradigm as the traditional approach used for comparison, (2) avoid the increased complexity of a more complicated classification scheme, and (3) achieved its goal of determining what the “main topics related to cyber security” are using machine learning.

B. *Pendar* (EX1005)⁴

Pendar refers to U.S. Patent No. 9,330,167 to Pendar.

Pendar discloses a system for classification and tagging of textual data. *Pendar* at abstract. It teaches that “each document in the corpus may be assigned one or more labels, and may be broken down into sentences so as to enable one or more queries of features.” *Pendar* at 7:64–8:1. “A subset of the one or more queries may then be selected based on their ability to identify a label associated to a particular document.” *Id.* at 8:1–3. *Pendar* allows for a hierarchy of labels, such as for example, “food and drink,” “seafood,” and “lobster.” *Id.* at 15:46–62.

⁴ EX2011 at ¶¶92–93 (*Pendar*).

C. *Frank* (EX1007)⁵

Frank refers to Eibe Frank, *Pruning Decision Trees and Lists*, University of Waikato, submitted January 2000.

Frank teaches “pruning algorithms for decision trees and lists that are based on significance tests.” *Frank* at abstract. *Frank* discusses various pruning methods, some of which rely on a “splitting criterion.” *Id.* at 21. *Frank* discloses several possible “splitting criterion.” *Id.* at 7 (Noting that “various split selection criteria have been designed, for example the information gain (Quinlan, 1986b), the Gini index (Breiman et al., 1984), and the gain ratio (Quinlan, 1992).”).

Notably, *Frank* discourages and teaches away from indiscriminate child-node pruning (which considers only the probability value of the parent tag, not those of the child tags to be pruned), because such “top-down pruning methods suffer from the ‘horizon effect’: a tree might be pruned even when it contains a subtree that ‘... would not have been pruned by the same criterion.’” *Frank* at 61 and 75. Further, *Frank* teaches away from using any single metric for pruning. *Frank* at 21 (“At each node, the learning algorithm selects the split that corresponds to the best value for the splitting criterion.”).

⁵ EX2011 at ¶¶94–96 (*Frank*).

D. *Ramírez-Corona* (EX1008)⁶

Ramírez-Corona refers to Mallinali Ramirez-Corona et al., *Multi-label Classification for Tree and Directed Acyclic Graphs Hierarchies*, Probabilistic Graphical Models, PGM 2014, (2014)

Ramírez-Corona discloses a Hierarchical Multi-label Classification (HMC) which assigns a set of classes to a single instance with the peculiarity that the classes are ordered in a predefined structure. *Ramírez-Corona* at abstract. It discloses various techniques for pruning tree structures.

E. *Chawla* (EX1009)⁷

Chawla refers to Nitesh V. Chawla et al., *SMOTE: Synthetic Minority Oversampling Technique*, Journal of Artificial Intelligence Research 16 (2002) 321–357.

Chawla addresses solutions to the class imbalance problem, in order to improve the accuracy of classifiers for a minority class in the training data.. *Chawla* at abstract. The class imbalance problem refers to classification categories that are not approximately equally represented. *Id.*

⁶ EX2011 at ¶¶97–98 (*Ramírez-Corona*).

⁷ EX2011 at ¶¶99–100 (*Chawla*).

F. *Xiao* (EX1010)⁸

Xiao refers to Tong Xiao et al., KNN and Re-ranking Models for English Patent Mining at NTCIR-7, proceedings of NTCIR-7 workshop meeting, (2008).

Xiao describes an English patent mining system which maps a research paper abstract into the IPC (International Patent Classification) taxonomy. *Xiao* at 333. Specifically, it “aims to assign one or multiple appropriate IPC codes to a non-patent document.” *Id.* *Xiao*’s input is the title and abstract of a published research paper. *Id.* The output is a list in which IPC codes are ranked in similarity scores. *Id.* *Xiao* uses various similarity calculation and ranking methods to improve its performance. *Id.* *Xiao* specifically notes that “if an IPC code is more strongly related to the input document, it should be assigned a larger score and ranked in an earlier position.” *Id.*

Notably, *Xiao* does not discuss using its method to modify the ground truth dataset for a machine learning algorithm.

G. *Ukrainczyk* (EX1011)⁹

Ukrainczyk refers to U.S. Patent No. 7,028,250 to Ukrainczyk et al.

Ukrainczyk discloses a method for automatically classifying text into categories. *Ukrainczyk* at abstract.

⁸ EX2011 at ¶¶101–103 (*Xiao*).

⁹ EX2011 at ¶¶104–105 (*Ukrainczyk*).

H. *Nunes* (EX1012)¹⁰

Nunes refers to Eric Nunes et al., *Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence*, Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics.

Nunes discloses a system for “cyber threat intelligence gathering from various social platforms on the Internet particularly sites on the darknet and deepnet.” *Nunes* at abstract. *Nunes* collects information “from hacker forum discussions and marketplaces offering products and services focusing on malicious hacking . . . for the purposes of identifying emerging cyber threats.” *Id.* Specifically, *Nunes* aims to identify “newly developed malware and exploits that have not yet been deployed in a cyber-attack.” *Id.*

Nunes discloses an “implementation and evaluation of learning models to separate relevant information from noise in the data collected from these online platforms.” *Id.* at 2. Notably, *Nunes*’ sole focus is on identifying products (from market data) and communications (from forum data) relevant to potential threats. *Id.*

¹⁰ EX2011 at ¶¶106–108 (*Nunes*).

V. THE '900 PATENT

A. Technical Background

Dr. Medvidovic provides a helpful explanation of the technical background of the '900 patent in paragraphs 69–74 of his declaration, reproduced in full here:

69. The '900 patent teaches various system and methods for the classification of data associated with the deep web or dark net. '900 patent at abstract. Enormous amounts of data relevant to potential criminal activity can be found on the dark web. “Considering the enormity of the data in those environments, there is an impending need to go beyond this broad categorization, thereby providing to security researchers a more granular, structural and interdependent classification of the available information.” *Id.* at 1:58–62.
70. The prior art suffered of complex problems. For example, “current technologies use learning models and techniques that do not address the issues of labeled data scarcity, nor do they address imbalanced data classes in the training set.” *Id.* at 2:10–12. “Training sets are labeled by hand, which is a time-consuming and typically an un-scalable process.” *Id.* at 2:14–15.
71. To address these shortcoming of the prior art, the '900 patent discloses a method for “automatically assigning one or more labels (tags), in a hierarchical structure, to discussion topics seen in d2web forums.” *Id.* at 3:12–15.
72. One embodiment of the '900 patent utilizes four main modules. *Id.* at 3:14–15. First, a specialized crawler that “crawls various d2web sites and extracts important information.” *Id.* at 3:16–18. Second, “data preprocessing and feature extraction is then performed over the forum discussion topics.” *Id.* at 3:24–26. Third, the system “extracts features from every topic on the forum.” *Id.* at 3:29–30. Fourth, “different machine classifiers are used to assign multiple tags to each topic, so that their performances can be compared.” *Id.* at 3:31–34.
73. Independent claim 1, for example, utilizes a ground truth set that is “generated from deep web forum information” and defines “a predetermined tag hierarchy including predetermined tag labels

corresponding [to] topics of the deep web forum information.” *Id.*, claim 1. The invention then accesses “data related to a discussion topic from a deep web forum” and extracts “a set of features” therefrom. *Id.* Then, a machine classifier is applied to the set of features which uses the ground truth to “generate a prediction list of tags for classifying the discussion topic” and “a prediction probability value for each tag.” *Id.* Finally, the invention adds all parent tags associated with each tag to the prediction list “based on a comparison of “the prediction probability value for the tag and a first predetermined threshold value.” *Id.*

74. In summary, the ’900 patent teaches a highly innovative method of solving the multiclass, multi-label classification problem for Classification of dark / deep web forum discussions, where the goal is to automatically assign one or more labels to each discussion. *Id.* at 39–42.

EX2011 at ¶¶69–74.

B. Prosecution History

The provisional application that resulted in the ’900 patent was filed on May 9, 2018, and the corresponding non-provisional application on May 7, 2019. EX1001. On February 24, 2021, an office action issued rejecting all claims. EX1002 at 219.

On August 13, 2021, patent owner amended the claims, adding the two claim limitations “wherein the prediction list includes a prediction probability value for each tag of the plurality of tags” (claim element 1.a.4) and “add all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value” (claim element 1.a.5). EX1002 at 284.

On November 12, 2021, a notice of allowance issued. EX1002 at 295. The examiner specifically noted that the prior art the prior art did not teach or suggest the amended claims. *Id.* at 301.

C. Priority Date

The '900 patent derives priority from provisional application No. 62/668,878, filed on May 9, 2018. The provisional application fully discloses the subject matter of the claims in the '900 patent. EX2011 at ¶76. Petitioner alleges the following priority dates for the asserted prior art:

- a. *Deliu* (EX1004): July 18, 2017 (Pet. at 9)
- b. *Pendar* (EX1005): May 13, 2013 (Pet. at 11)
- c. *Frank* (EX1007): January 8, 2001 (Pet. at 12)
- d. *Ramírez-Corona* (EX1008): 2014 (Pet. at 13)
- e. *Chawla* (EX1009): June 2002 (Pet. at 14)
- f. *Xiao* (EX1010): December 2008 (Pet. at 15)
- g. *Ukrainczyk* (EX1011): May 25, 2001 (Pet. at 15)
- h. *Nunes* (EX1012): November 8, 2016 (Pet. at 16)

Since these dates predate the provisional patent application as well, Patent Owner's analysis relies on the specification of the '900 patent itself. However, Patent Owner is still investigating the veracity of Petitioner's alleged dates. To the extent these dates are inaccurate and in fact post-date the provisional application,

Patent Owner reserves the right to identify the respective portions of the provisional application supporting the claims of the '900 patent.

VI. LEVEL OF ORDINARY SKILL IN THE ART

In consideration of the type of problems encountered in the art, the prior solutions to those problems found in the prior art references, the rapidity with which innovations are made, the sophistication of the technology, the level of education of active workers in the field, and in reliance on Dr. Medvidovic's expert opinion, Patent Owner defines the relevant skill in the art as follows: A person of ordinary skill in the art ("POSITA") at the time of the filing of the '900 patent would have at least a Bachelor's degree in computer science or computer engineering, and 2–4 years of experience working in the field of software development, distributed systems, data mining, and cybersecurity. A graduate degree in computer science, computer engineering, artificial intelligence, cybersecurity, or software engineering may be substituted wholly or in part for work experience. EX2011 at ¶37. Dr. Medvidovic qualifies as a POSITA. *Id.* at ¶38.

VII. PATENT OWNER'S CLAIM CONSTRUCTIONS

Patent claims are construed using the framework set forth in *Phillips v. AWH Corp.*, 415 F.3d 1303, 1312–15 (Fed. Cir. 2005) (*en banc*). CrowdStrike relies on the plain and ordinary meaning of all claim terms. Pet. at 5. Patent Owner applies

the ordinary meaning of the terms as a POSITA would have understood them at the time of the invention.

Petitioner notes two typographical errors in claims 3 and 4. Pet. at 6–9. Regarding claim 3, Petitioner identifies the following correction: “wherein the processor is configured to remove all child tags of the plurality of tags corresponding to a parent tag from the prediction list of tags where a probability value associated with the ~~child~~ [parent] tag is below a second predetermined threshold value.”

Similarly, for claim 4, Petitioner identifies the following correction: “wherein the processor is configured to add all parent tags of the plurality of tags corresponding to a child tag if a probability value associated with the ~~parent~~ [child] tag is above the first predetermined threshold value.”

Patent Owner agrees with these typographical corrections.

VIII. PETITIONER HAS FAILED TO DEMONSTRATE A REASONABLE LIKELIHOOD OF PREVAILING AS TO ANY CHALLENGED CLAIM.

A. GROUND 1: Obviousness in Light of *Deliu* and *Pendar*

Petitioner asserts that claims 1, 2, 4, 5, 10, 12, and 13 are obvious in light of *Deliu* and *Pendar*.

1. Claim 1

Claim 1. A computer-implemented system for improving classification of criminal activities arising from cyber environments, comprising:

a) Claim 1.a.1 access a ground truth dataset generated from deep web forum information, the ground truth dataset defining a predetermined tag hierarchy including predetermined tag labels corresponding [to] topics of the deep web forum information;

Petitioner argues that “*Deliu* tested its machine learning classification methods ‘on a real hacker forum.’” Pet. at 18. This is not correct. EX2011 at ¶116. *Deliu* tested its experiments “on a leaked dataset generated from the Nulled.IO dark-web forum.” Petition at 18 (“*Deliu* obtained the forum information from a whistleblower website.”) Although “Nulled.IO forum was originally part of the dark web” (*Id.*), the information in that forum was extracted from the dark web forum and heavily processed before *Deliu* retrieved it. EX2011 at ¶116 (citing *Deliu* at 62 (“The data of this forum already exists in a mysql format and no special web crawlers are needed to collect it.”)) Indeed, “the complete relational database of the forum has been leaked which simplifies the process of extracting the field of interest from the forum.” *Deliu* at 73–74. The ’900 patent, on the other hand, explicitly “crawls various d2web sites and extracts important information from html pages, storing it in a database.” ’900 patent at 3:17–19. “Specialized crawlers automatically connect and authenticate the d2web sites through Tor.” *Id.* at 3:19–21.

Further, *Deliu* does not use the term “ground truth.” Petitioner relies on *Deliu*’s disclosure that “in order to build a model which is able to filter out posts which are not relevant to security, the presence of labeled samples is required...we

constructed a dataset by manually labeling some of the posts in the identified forum.”
Petition at 19.

Petitioner admits that “*Deliu* does not specifically discuss that its multi-class dataset includes hierarchical data such that it defines a “predetermined tag hierarchy,”” and cites *Pendar* for a “predefined hierarchical structure.” Pet. at 21 (citing *Pendar* at 15:46-62). One such example is the label “burgers” being associated with “traditional American cuisine” which in turn is associated with “food and drink.” *Id.*

Petitioner’s argument fails because a POSITA would not have been motivated to hierarchically label *Deliu*’s multiclass dataset pursuant to *Pendar*. EX2011 at ¶119. *Deliu*’s explicit goal is to answer two research questions: “1. How can different Machine Learning methods be used to classify the contents of hacker forums?” and “2. What are the main topics related to cyber security on such forums, and how does filtering through classification affect the discovered topics?” *Deliu* at 19. “Neither of these goals would have benefitted from *Pendar*’s hierarchical structure.” EX2011 at ¶119. Because *Pendar*’s hierarchical labeling does not advance either of these expressly stated research questions, and would only add complexity to *Deliu*’s analysis, a POSITA would have had no reason to modify *Deliu*’s binary- and multi-class labeling scheme to conform to *Pendar*’s hierarchy.

This is explained in more detail in the motivation to combine section below. *See* Subsection 8 (“Motivation To Combine”).

Petitioner’s argument that *Deliu* “envisions utilizing other deep learning algorithms to compare to the algorithms tested in its experiments” (Pet. at 22) does not support the conclusion that a POSITA would have been motivated to implement a hierarchical classification structure. “To the contrary, a meaningful comparison would require that the classification paradigm remain the same, i.e., either binary or multi-class as in the traditional approaches *Deliu* used for comparison.” EX2011 at ¶120. A different approach would have made *Deliu*’s stated goal of “compar[ing] to the algorithms tested in its experiments” impossible, or at a minimum, would have required significant non-trivial effort to build a comparison scheme. *Id.* at ¶120 (this effort would have been beyond the skill of a POSITA). “There is no teaching in *Deliu* that supports a motivation or interest to explore other classification paradigms.” *Id.*

Petitioner further cites *Hoffman*’s observation that “hierarchical SVM outperforms flat SVM.” Pet. at 22 (citing EX1015 (“*Hoffman*”) at 9). However, “*Hoffman*’s analysis relates to a very limited use case of analyzing WIPO (World Intellectual Property Organization) patent documents which are classified in a four level hierarchy, i.e., sections, classes, subclasses and groups, according to the International Patent Classification standard.” EX2011 at ¶121 (citing *Hoffman* at 9).

Dr. Medvidovic explains by analogy: when comparing the performance of several automobile engines in different kinds of weather, the argument that jet engines perform better in cold weather is an unhelpful non sequitur. *Id.* at ¶121. The performance of hierarchical SVM (Support Vector Machine) generally, or in *Hoffman*'s particular use case, does not support modifying *Deliu* to utilize a different classification paradigm, because "that would lead to comparison of two different classification schemes to one another." *Id.* at ¶121.

In *Hoffman*'s setting, the classifier is designed specifically for that four-level hierarchy and assumes that documents are assigned within that fixed multi-level scheme. *Id.* at ¶121. By contrast, *Deliu*'s experiments are conducted on flat binary and multi-class datasets whose labels are not organized into such a hierarchy. *Id.* at ¶119. A POSITA would understand that changing *Deliu*'s classifier to a hierarchical SVM of the type evaluated in *Hoffman* would require restructuring the way labels are defined and applied, so that the resulting classification scheme is no longer the same as the flat schemes used in *Deliu*'s comparisons. *Id.* at ¶121. Therefore, the performance of *Hoffman*'s hierarchical SVM could not be directly compared, on an equal footing, with the traditional flat SVM and other classifiers that *Deliu* actually evaluated for its research questions. *See Id.* Thus, a POSITA would remain focused on *Deliu*'s stated goal of comparing alternative machine learning methods on the

same hacker-forum classification task and would see no motivation to replace *Deliu*'s flat scheme with *Hoffman*'s hierarchical approach. *Id.*

The same is true for *Deliu*'s other classification techniques, because *Deliu* would have avoided using one that changes the paradigm so dramatically as to make the comparison useless. *Id.* Thus, in view of *Deliu*'s stated goal of comparing different machine learning methods on the same binary and multi-class tasks, a POSITA would have recognized that moving to a hierarchical scheme of the type analyzed in *Hoffman* would defeat the purpose of *Deliu*'s comparison, instead of advancing it.

Finally, especially given Petitioner's statement that "classifying ground truth data into a hierarchical form was a well-known strategy in machine learning," the fact that *Deliu* intentionally chose and disclosed only a binary- and multi-class labeling scheme, without suggesting any hierarchical extension, further supports that modifying the underlying label structure to be hierarchical was not a logical extension of *Deliu*'s work. Pet. at 23.

b) Claim 1.a.4 apply a machine classifier to the set of features to generate a prediction list of tags for classifying the discussion topic, wherein the prediction list includes a prediction probability value for each tag of the plurality of tags; and

Petitioner admits that *Deliu* does not teach a "prediction probability value" for each tag. Pet. at 26. *Pendar* does not either.

Petitioner relies on *Pendar*'s teaching that "a label is applied by summing the scores of each positive query such that when a document receives a summed score above a predetermined threshold, the label is assigned." Pet. at 27 (citing *Pendar* at 14:37–67 and 9:56-62, and EX1003 at ¶¶125–126). Petitioner argues that the "summed scores for each label are 'values' that are compared to a threshold to predict a label and are therefore prediction probability values as claimed." *Id.* This is incorrect.

As explained by Dr. Medvidovic, in *Pendar*, the summed value identified by Petitioner is used "to determine a label for the document." EX2011 at ¶123 (citing *Pendar* at 10:6–7; *see also id.* at 14:63–67 ("if a document retrieved from an electronic source has summed scores exceeding a pre defined threshold for 'pizza' and 'spaghetti,' each label may be associated with the document and thus associated with the particular merchant or service provider.")). *Pendar* uses its summed value "to determine whether or not to apply a label to a document in the first place." EX2011 at ¶123.

In stark contrast, "the probability value in the '900 patent does not related to whether or not a label is applied to the discussion topic," in the first place. *Id.* at ¶124. Claim 1.a.4 requires that the prediction list include "a prediction probability value for each tag of the plurality of tags," and claim 1.a.5 then recites using those same probability values to adjust the hierarchical tag structure. Dr. Medvidovic

explains, only “after a label is applied, [the] probability value is determined to allow the system to ‘regulate[] the way parent tags are added and child tags are removed, based on the probability of the predicted tag.’” *Id.* at ¶124 (citing ’900 patent at 6:61–63). Indeed, the very next claim element (1.a.5) relies on this probability value to evaluate whether all parent tags associated with the given tag should be added to the prediction list. EX2011 at ¶124. This is based on whether the prediction probability value for the given tag reaches the predetermined threshold value. *Id.* (citing ’900 patent, claim element 1.a.5). This is further confirmed by the specification. ’900 patent at 11:30–32 (“parent tags were added and child tags removed based on the probability of the predicted tag”). Unlike in *Pendar*, this analysis does not affect whether or not the given tag itself—as opposed to its parent tags—are added to the prediction list.

The ’900 patent further explains the purpose of this functionality, which is to increase the accuracy of the system. ’900 patent at 11:36–38 (“Removing child tags reduces false positives, increasing the precision score, while adding parent tags reduces false negatives, increasing the recall score.”); *see also* Fig. 10 and 11:40–44 (“with an increase in “add parent threshold” (α) value, the recall score increased and precision score decreased. Conversely, with an increase in “remove child threshold” value (β), the precision score increased and recall score decreased.”);

'900 patent at 11:47–49 (disclosing the optimal configuration as “ $\beta=0.5$ and $\alpha=0.9$.”); EX2011 at ¶125.

Dr. Medvidovic further explains this functionality by reference to Algorithm 1 of the '900 patent.

Algorithm 1 Combination of Adding Parent Tags and Removing
Child Tags (T)

Input: Prediction list of tags (T) for each document.
Output: Prediction list of tags that respect the tag hierarchy for each document.
1: $\beta \leftarrow$ remove child threshold value
2: $\alpha \leftarrow$ add parent threshold value
3: for $t := 1$ to T do
4: $p(t) \leftarrow$ prediction probability of tag t.
5: if $p(t) < \beta$ then remove all child tags of tag t from prediction list
6: for $t := 1$ to T do
7: $p(t) \leftarrow$ prediction probability of tag t.
8: if $p(t) > \alpha$ then add all parent tags of tag t from prediction list
return updated prediction list

EX2011 at ¶126 (citing '900 patent at 6:45-56 (Algorithm 1)). This algorithm confirms that “the probability value $p(t)$ relates to tags that are already in the prediction set.” *Id.* “The algorithm then compares each tag’s probability value to a threshold, i.e., α as ‘add parent threshold’ and β as ‘remove child threshold.’ *Id.* *Deliu* and *Pendar* do not disclose such a probability value for tags that have already been added to the prediction list (i.e., assigned to a topic in *Pendar*).

The prosecution history corroborates this analysis. The “probability value” claim element was specifically added during prosecution to address an initial rejection. EX1002 at 284 (adding claim element “wherein the prediction list includes a prediction probability value for each tag of the plurality of tags”). In

reliance on this new element, the notice of allowance explicitly noted that the prior art does not disclose the amended claims. *Id.* at 301.

Further, a POSITA would not have been motivated to combine *Deliu* with *Pendar* (see section 8 below). Petitioner argues that “a POSITA would have been motivated to modify *Deliu*’s classifiers to implement *Pendar*’s multi-label querying techniques to facilitate assigning more than one label to each document.” Pet. at 27. But, as Dr. Medvidovic explains, “the cited section of *Deliu* (noting that ‘classifier performance starts to get better than traditional approaches only when the dataset contains millions of samples’) does not support this.” EX2011 at ¶128 (citing EX1016 (*Deliu* paper) at 7–8). Neither Petitioner nor *Deliu* provide an explanation as to why increasing the size of the dataset provides motivation to assign more than one label to each document or to use a hierarchy of labels. The passage Petitioner cites from *Deliu* merely observes that certain classifier performance characteristics improve when the dataset contains millions of samples, and says nothing about changing the underlying labeling scheme or moving from single-label to multi-label or hierarchical classification. EX2011 at ¶128. As Dr. Medvidovic explains, *Deliu* frames its study around comparing the performance of different classifiers on its chosen binary- and multi-class datasets, and nowhere suggests that altering those labeling assumptions to a multi-label or hierarchical structure would advance its stated research goals. *Id.* at ¶¶120 and 128. Further, Petitioner’s statement that

“[c]omplex topologies would benefit from classifying a document into multiple categories to account for overlap where an item can belong to several categories at once, providing a more realistic classification for the dataset that accounts for ambiguity in the data” is conclusory and provides no explanation for its argument. Pet. at 27. Petitioner cites no support in either *Deliu* or *Pendar* to support this naked assertion. The same is true for Petitioner’s statement that “*Pendar*’s predicted labels and scores would have provided important information regarding the ‘informativeness’ of each label.” Pet. at 27–28. The cited references provide no support for this argument. EX2011 at ¶128.

c) Claim 1.a.5 add all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value.

Petitioner argues that “*Deliu-Pendar* assigns labels based on its classifiers comparing scores to thresholds.” Pet. at 30 (citing *Pendar* at 14:48-51 (checking the summed score to a threshold score) and 14:51-57 (if the summed score exceeds a predetermined threshold, then the label is applied to the document)). As noted regarding claim element 1.a.4, *Pendar*’s “summed score” is used “to determine a label for the document” itself. *Pendar* at 10:6–7 and 14:63–67. In other words, *Pendar*’s summed value determines whether or not to apply a label to a document in the first place; it does not relate to parsing the entire prediction list afterwards to determine whether parent labels should be added as claimed. EX2011 at ¶129.

In stark contrast, the probability value in the '900 patent does not determine whether or not the predicted tag is added. EX2011 at ¶130. Instead, “a probability value is assigned to each predicted tag.” *Id.* (citing '900 patent at 6:61–63). “That probability value then allows the system to ‘regulate[] the way parent tags are added and child tags are removed, based on the probability of the predicted tag.’” *Id.* (further citing '900 patent at 11:30–32 (“parent tags were added and child tags removed based on the probability of the predicted tag”)).

Dr. Medvidovic specifically notes that “this functionality increases the accuracy of the system.” EX2011 at ¶131 (citing '900 patent at 11:36–38; Fig. 10 and 11:40–44 (“with an increase in “add parent threshold” (α) value, the recall score increased and precision score decreased. Conversely, with an increase in “remove child threshold” value (β), the precision score increased and recall score decreased.”); and '900 patent at 11:47–49).

Dr. Medvidovic provides further explanation by reference to Algorithm 1:

Algorithm 1 Combination of Adding Parent Tags and Removing
Child Tags (T)

Input: Prediction list of tags (T) for each document.
Output: Prediction list of tags that respect the tag hierarchy for each document.

- 1: $\beta \leftarrow$ remove child threshold value
- 2: $\alpha \leftarrow$ add parent threshold value
- 3: for $t := 1$ to T do
- 4: $p(t) \leftarrow$ prediction probability of tag t.
- 5: if $p(t) < \beta$ then remove all child tags of tag t from prediction list
- 6: for $t := 1$ to T do
- 7: $p(t) \leftarrow$ prediction probability of tag t.
- 8: if $p(t) > \alpha$ then add all parent tags of tag t from prediction list

return updated prediction list

'900 patent at 6:45-56 (Algorithm 1). “The algorithm explicitly confirms that every tag ‘t’ in the prediction set has a probability value ‘p(t)’ which is compared to a threshold, i.e., α as ‘add parent threshold’ and β as ‘remove child threshold.’” EX2011 at ¶132. “Neither *Deliu* nor *Pendar* disclose such a probability value for tags that have already been assigned.” *Id.*

Petitioner argues that *Deliu-Pendar* teaches adding all dominant labels associated with a label to the prediction list for labels that exceeded a predetermined threshold. Pet. at 30–31. But this mischaracterizes the combination. As Dr. Medvidovic explains:

In Petitioner’s *Deliu-Pendar*, if the summed score exceeds a threshold, the given label is added. But, unlike the '900 patent, dominant (parent) labels are also added. In contrast, this claim element requires that for each label “t”, the parent label is added if and only if a condition precedent is satisfied, i.e., that the probability value for tahe “t” exceeds a threshold value. This is fundamentally different from *Deliu-Pendar*.

EX2011 at ¶133. In other words, *Deliu-Pendar* adds parent labels automatically whenever the child label’s score exceeds its threshold, whereas claim 1.a.5 requires a separate comparison between the prediction probability value for the tag and a threshold before adding its parent tags.

Dr. Medvidovic further notes that “the '900 patent discloses—but does not claim—the *Deliu-Pendar* algorithm, called “adding parent tag.” EX2011 at ¶134 (citing '900 patent at 6:28–35 (“If a child tag is predicted by the classifier and if the corresponding parent tags in the hierarchy are not predicted, then all the parent tags

are added to the prediction list.”)). Conversely, the ’900 patent “also discloses but does not claim an algorithm called ‘removing child tag.’” EX2011 at ¶134 (citing ’900 patent at 6:36–43). These two unclaimed procedures add and remove parent and child tags whenever the corresponding child or parent tag is predicted, without any probability-based control. *Id.* By contrast, the claimed method introduces prediction probability values and thresholds to fine-tune when parent tags are added and child tags are removed. The ’900 patent shows that suitable choices of α and β improve performance. ’900 patent at 11:47–49 (“The best F1 score was achieved when $\beta=0.5$ and $\alpha=0.9$. The corresponding F1 score is 31.9%, precision score is 73.43% and recall score is 20.4%.”).

The prosecution history confirms this distinction. The claims were amended with the claim element “add all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value.” EX1002 at 284. The notice of allowance explicitly confirms that the prior art does not disclose this element. *Id.* at 301.

Further, a POSITA would not have been motivated to combine *Deliu* with *Pendar* as explained in detail regarding claim element 1.a.4 and in section 8.

2. Claim 2

Claim 2. The computer-implemented system of claim 1, wherein a predetermined tag hierarchy is divided into parent tags and child tags,

wherein a topic represented by a child tag is representative of a subset of a topic represented by a parent tag, wherein each child tag corresponds to a particular parent tag, and each tag is assigned a corresponding probability value.

Claim 2 depends on claim 1 and is not disclosed or suggested for the reasons set forth regarding claim 1.

Petitioner references its argument regarding claim element 1.a.4 for the claim element “each tag is assigned a corresponding probability value.” Pet. at 34. Patent Owner therefore incorporates its response thereto and to related claim element 1.a.5. This claim element is not disclosed for the reasons set forth there.

3. Claim 4

Claim 4. The computer-implemented system of claim 1, wherein the processor is configured to add all parent tags of the plurality of tags corresponding to a child tag if a probability value associated with the ~~parent~~ [child] tag is above the first predetermined threshold value.¹¹

Claim 4 depends on claim 1 and is not disclosed or suggested for the reasons set forth regarding claim 1.

Petitioner references its argument regarding claim element 1.a.5 for this claim. Pet. at 35. Patent Owner therefore incorporates its response thereto. This claim is not disclosed for the reasons set forth there.

¹¹ Claim 4 has been edited to correct the indicated typographical error.

4. Claim 5

Claim 5. The computer-implemented system of claim 1, wherein the set of features are extracted using textual feature extraction to identify tags associated with topic titles of the data.

Claim 5 depends on claim 1 and is not disclosed or suggested for the reasons set forth regarding claim 1.

5. Claim 10

Claim 10. The computer-implemented system of claim 1, wherein the set of features are extracted from the data using a neural network, wherein the neural network is configured to learn a continuous distributed vector representation for documents and words within the data.

Claim 10 depends on claim 1 and is not disclosed or suggested for the reasons set forth regarding claim 1.

6. Claim 12

Claim 12. A method, comprising:

a.3 apply a machine classifier to the set of features to generate a prediction list of tags for classifying the topic, wherein the prediction list includes a prediction probability value for each tag of the plurality of tags; and

Petitioner references its argument regarding claim element 1.a.4 for this claim.

Pet. at 40. Patent Owner therefore incorporates its response thereto. This claim is not disclosed for the reasons set forth there.

a.4 adding all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value.

Petitioner references its argument regarding claim element 1.a.5 for this claim. Pet. at 40. Patent Owner therefore incorporates its response thereto. This claim is not disclosed for the reasons set forth there.

7. Claim 13

Claim 13. The method of claim 12, further comprising: outputting an update to the prediction list of tags, a total number of the tags of the prediction list of tags being different from the update to the prediction list of tags to maintain a predetermined tag hierarchy.

Claim 13 depends on claim 12 and is not disclosed or suggested for the reasons set forth regarding claim 12.

“In the context of claims 12 and 13, the term ‘outputting an update to the prediction list of tags’ refers to the algorithm ‘Combination of Adding Parent Tag and Removing Child Tag’ disclosed in the ’900 patent.” EX2011 at ¶144 (citing ’900 patent at 6:59–67). The output of this algorithm (highlighted below) is an “updated prediction list.”

Algorithm 1 Combination of Adding Parent Tags and Removing
Child Tags (T)

Input: Prediction list of tags (T) for each document.
Output: Prediction list of tags that respect the tag hierarchy for each document.

- 1: $\beta \leftarrow$ remove child threshold value
- 2: $\alpha \leftarrow$ add parent threshold value
- 3: for $t = 1$ to T do
- 4: $p(t) \leftarrow$ prediction probability of tag t.
- 5: if $p(t) < \beta$ then remove all child tags of tag t from prediction list
- 6: for $t = 1$ to T do
- 7: $p(t) \leftarrow$ prediction probability of tag t.
- 8: if $p(t) > \alpha$ then add all parent tags of tag t from prediction list

return updated prediction list

'900 patent at 6:45-56 (Algorithm 1). Dr. Medvidovic explains the functionality:

The algorithm compares the probability value of each tag “t” in the prediction list to a threshold, i.e., α as “add parent threshold” and β as “remove child threshold.” If $p(t) > \alpha$, all parent tags of tag “t” are added. If $p(t) < \beta$, all child tags of tag “t” are removed. If neither comparison is satisfied, then no action is taken for that tag “t.” When done, the updated prediction list is returned. Claim 13 claims the likely instance that this updated prediction list differs in length from the original prediction list due to the added parent tags and/or removed child tags.

EX2011 at ¶145 (citations removed). “Neither *Deliu* nor *Pendar* disclose such a probability value for each predicted tag, not such an operation that would result in a modified prediction list.” EX2011 at ¶146.

Petitioner mischaracterizes *Deliu* and *Pendar*, and cites the automatic addition of parent tags taught in *Pendar* for this claim element. Pet. at 41. This is wrong. Unlike the '900 patent, “*Pendar* teaches that—if a set threshold is reached—the label is applied and all parent labels are automatically applied as well.” EX2011 at ¶147 (emphasis added). “This is done in the first instance, i.e., in the original prediction list returned by *Pendar*.” *Id.* Therefore, *Pendar* does not disclose an original list, nor any kind of processing of that original list (by either adding parent tags or removing child tags) that might result in an updated prediction list of a different length. *Id.*

8. Motivation To Combine

A POSITA would not be motivated to combine *Deliu* with *Pendar*. *Deliu* compares “various machine learning approaches to identify the main topics in hacker forums.” EX2011 at ¶148. “*Pendar* teaches a multilevel document classification scheme.” *Id.*

As Dr. Medvidovic explains, a POSITA would not have been motivated to modify *Deliu* to hierarchically label the multiclass dataset pursuant to *Pendar*’s label hierarchy. EX2011 at ¶149. To the contrary, *Deliu* identifies its dual goals as “1. How can different Machine Learning methods be used to classify the contents of hacker forums?” and “2. What are the main topics related to cyber security on such forums, and how does filtering through classification affect the discovered topics?” *Deliu* at 19. “Neither of these goals would have benefitted from *Pendar*’s hierarchical structure.” EX2011 at ¶149.

Regarding the first goal, *Deliu* states that the “classification performance of recent Deep Learning algorithms is compared with more traditional Machine Learning methods in a both a binary- and multi-class dataset constructed from a real hacker forum.” *Deliu* at 19. “To the best of our knowledge, this comparison is a novel contribution to the cyber security community.” *Id.* Thus, “*Deliu* focuses on the efficiency of the “how” to classify, as opposed to the final structure of the

classification itself, i.e., whether it is hierarchical or not.” EX2011 at ¶150. *Deliu’s* conclusion to research question 1 confirms this:

After the focus deep learning has gained recently, we compare its performance to more traditional algorithms. Based on the success it has shown in similar problems, we build the hypothesis that the deep ConvNN used in this thesis will outperform other conventional classifiers. The results of the experiments show that the traditional SVM classifier with character n-gram features perform at least as good as the ConvNN. Therefore, our hypothesis is not supported.

Deliu at 103. *Deliu* thus only compared the performance of various machine learning approaches with each other. EX2011 at ¶150. “*Deliu* identifies no motivation to increase the complexity of the classification scheme (e.g., to a hierarchical scheme).” *Id.* To the contrary, *Deliu* explicitly teaches that SVM’s are preferred. *Id.* (citing EX1016 at 8 (“Given the computational complexity of CNN architectures, our results indicate that SVMs are superior for the purposes of practical, real-time CTI applications.”)).

To the contrary, Dr. Medvidovic explains that “‘*traditional SVM*’ is a binary classification model.” EX2011 at ¶151 (citing *Deliu* at 55 (*Nunes*’ uses “binary classification to categorize the hacker forums posts and marketplace products into two classes: (a) relevant and (b) irrelevant to cyber security.”); and 104 (“*Nunes et al.* used *traditional SVM* classifier with n-grams as features in a similar dataset.”)) (emphasis added). “Thus, *Deliu* was motivated to use the same binary classification model to be able to compare equal paradigms.” *Id.*

The second goal seeks to identify the “main topics related to cyber security” on such forums. “This goal explicitly focuses on “main topics” and would therefore not benefit from a determination of “sub-topics” in a topic hierarchy.” EX2011 at ¶152.

Further, Dr. Medvidovic notes that “*Deliu* confirms that multi-class data sets increase the complexity of the problem.” EX2011 at ¶153 (citing *Deliu* at 106 (“Manual labeling might be resource demanding especially in the case of the multi-class dataset”); *Id.* at 63 (“We must emphasize that this kind of [multi-class] labeling is much more demanding.”); *Id.* at 74 (“the construction of the [multi-class] dataset is much more demanding.”)). Notably, “imposing a hierarchical classification structure is non-trivial and further increases the complexity of the problem identified by *Deliu*.” *Id.*

Further, Petitioner’s combination, which would introducing a hierarchical classification scheme, “would have required developing a whole separate comparison framework in order to make it work with *Deliu*’s classification.” EX2011 at ¶155. Petitioner ignores that this “would require significant additional experimentation and would have been outside the skill set of a POSITA.” *Id.*

Therefore, “*Deliu* teaches away from using a hierarchical classification scheme because it seeks to (1) use the same paradigm as the traditional approach used for comparison, (2) avoid the increased complexity of a more complicated

classification scheme, and (3) achieved its goal of determining what the ‘main topics related to cyber security’ are using machine learning.” *Id.* at ¶154.

B. Ground 2: Obviousness in Light of *Deliu*, *Pendar*, and *Frank*

Petitioner asserts that claim 3 is obvious in light of *Deliu*, *Pendar*, and *Frank*.

1. Claim 3

Claim 3. The computer-implemented system of claim 1, wherein the processor is configured to remove all child tags of the plurality of tags corresponding to a parent tag from the prediction list of tags where a probability value associated with the ~~child~~ [parent] tag is below a second predetermined threshold value.¹²

Claim 3 depends on claim 1 and is not disclosed or suggested for the reasons set forth regarding claim 1.

Petitioner references its argument regarding claim element 1.a.4 for this claim. Pet. at 44. Patent Owner therefore incorporates its response thereto. This claim is not disclosed for the reasons set forth there.

Petitioner identifies *Frank*’s “splitting criterion” as the probability value and argues that “if the splitting criterion is below the threshold (i.e., below a second predetermined threshold), then the subtree is removed (i.e., remove all child tags) in place of a single leaf node.” Pet. at 44. This functionality does not teach or suggest claim 3. As noted by Dr. Medvidovic, “Petitioner mischaracterizes this functionality

¹² Claim 3 has been edited to correct the indicated typographical error.

by omitting that an additional requirement is imposed by *Frank* that prevents the removal of the child tags.” EX2011 at ¶159 (citing *Frank* at 62 (“However, one additional constraint is imposed: if the subtree contains at least one node whose value is greater than the threshold, it will not be pruned.” “This means that a subtree is only considered for pruning if all its successors are leaf nodes.”) Thus, *Frank* does not disclose the claimed functionality of “removing all child tags” if the “splitting criterion” is below a threshold. Instead, “*Frank* only removes the child nodes if the splitting criterion of the node itself and the splitting criterion of every child node fall below the set threshold.” EX2011 at ¶159. This is a significant difference to claim 3, which requires removal of all child tags based only on the probability value of the parent tag. *Id.*

Further, Dr. Medvidovic explains that “*Frank*’s ‘splitting criterion’ is not a ‘probability value.’” EX2011 at ¶160. *Frank* discloses various options as possible “splitting criteria,” but none are a “probability value.” *Frank* at 7 (noting only that “various split selection criteria have been designed, for example the information gain (Quinlan, 1986b), the Gini index (Breiman et al., 1984), and the gain ratio (Quinlan, 1992).”). “None of these potential split selection criteria are a ‘probability value.’” *Id.* Petitioner further identifies *Pendar*’s “summing scores” as a probability value, but for the reasons stated regarding claim element 1.a.4, that also does not disclose the claimed probability value. *Id.*

2. Motivation To Combine

A POSITA would not have been motivated to combine *Deliu* and *Pendar* for the reasons stated regarding ground 1 above. Further, a POSITA would not have been motivated to combine these two references with *Frank*. “*Frank* actually discourages and teaches away from the invention in claim 3 (which considers only the probability value of the parent tag, not those of the child tags to be pruned), because such ‘top-down pruning methods suffer from the ‘horizon effect’: a tree might be pruned even when it contains a subtree that ‘... would not have been pruned by the same criterion.’” EX2011 at ¶161 (citing *Frank* at 61 and 75). This disclosure in *Frank* would have motivated a POSITA to avoid what is taught in claim 3, as it could have the effect of removing child tags that have probability values that would not have resulted in removal by the same threshold, i.e., to avoid the ‘horizon effect.’ *Frank* also teaches away from using a single metric, such as the claimed “probability value.” EX2011 at ¶161 (citing *Frank* at 21 (“At each node, the learning algorithm selects the split that corresponds to the best value for the splitting criterion.”)).

C. Ground 3: Obviousness in Light of *Deliu*, *Pendar*, and *Ramírez-Corona*

Petitioner asserts that claim 3 is obvious in light of *Deliu*, *Pendar*, and *Ramírez-Corona*. However, Petitioner notes that “Ground 3 challenges Claim 3 uncorrected.” Pet. at 43 (fn. 4) (emphasis added). Since both parties agree that

Claim 3 is properly interpreted as corrected (EX2011 at ¶162), shown below, ground 3 is moot.

Claim 3. The computer-implemented system of claim 1, wherein the processor is configured to remove all child tags of the plurality of tags corresponding to a parent tag from the prediction list of tags where a probability value associated with the ~~child~~ [parent] tag is below a second predetermined threshold value.¹³

D. Ground 4: Obviousness in Light of *Deliu*, *Pendar*, and *Chawla*

Petitioner asserts that claims 6 and 14 are obvious in light of *Deliu*, *Pendar*, and *Chawla*.

1. Claim 6

Claim 6. The computer-implemented system of claim 1, wherein a set of synthetic data samples are created for addition to the ground truth dataset to address class imbalance by creating a set of synthetic sample points along a line defined between a first minority feature vector and a second feature vector and adding the set of synthetic sample points to the ground truth dataset.

Claim 6 depends on claim 1 and is not disclosed or suggested for the reasons set forth regarding claim 1.

Petitioner references its argument regarding claim 1 for this claim. Pet. at 48. Patent Owner therefore incorporates its response thereto. This claim is not disclosed for the reasons set forth there.

2. Claim 14

Claim 14. The method of claim 12, further comprising generating a predetermined tag hierarchy based on a ground truth dataset, and leveraging

¹³ Claim 3 has been edited to correct the indicated typographical error.

semi-supervised learning to address class imbalance and increase accuracy of the machine classifier.

Claim 14 depends on claim 12 and is not disclosed or suggested for the reasons set forth regarding claim 12.

Petitioner references its argument regarding claim element 1.a.1 for this claim. Pet. at 51. Patent Owner therefore incorporates its response thereto. This claim is not disclosed for the reasons set forth there.

E. Ground 5: Obviousness in Light of *Deliu*, *Pendar*, and *Xiao*

Petitioner asserts that claim 7 is obvious in light of *Deliu*, *Pendar*, and *Xiao*.

1. Claim 7

Claim 7. The computer-implemented system of claim 1, wherein the ground truth dataset is supplemented by, for every sample of the ground truth dataset in a minority class, adding top similar documents from the deep web forum information, as identified using an elastic search similarity score, and adding them to the ground truth dataset.

Claim 7 depends on claim 1 and is not disclosed or suggested for the reasons set forth regarding claim 1.

Petitioner references its argument regarding claim 1 for this claim. Pet. at 52. Patent Owner therefore incorporates its response thereto. This claim is not disclosed for the reasons set forth there.

Xiao focuses on mapping research paper abstracts into the International Patent Classification (IPC) taxonomy. *Xiao* at 333. Specifically, it “aims to assign one or multiple appropriate IPC codes to a non-patent document.” *Id.* *Xiao*’s input is the

title and abstract of a published research paper. *Id.* The output is a list in which IPC codes are ranked in similarity scores. *Id.* *Xiao* specifically notes that “if an IPC code is more strongly related to the input document, it should be assigned a larger score and ranked in an earlier position.” *Id.* “However, *Xiao* does not modify the ground truth dataset for a machine learning algorithm.” EX2011 at ¶171.

2. Motivation To Combine

As Dr. Medvidovic explains, *Xiao* specifically identifies the “huge amount of training samples” as a “challenging problem.” EX2011 at ¶172 (citing *Xiao* at 335):

[T]he number of patent samples for training is over 3 million. In real-world applications, how to train a supervised classifier on such a large scale of data set is a critical issue. For example, training SVM-based classifier on such large corpus is also intractable due to the high computation cost.

Id. This is the exact opposite of the problem described by *Deliu*. EX2011 at ¶172. (citing Pet. at 52 (“*Deliu* teaches ‘snowball sampling[,]’ ‘a technique used to increase the number of samples [i.e., adding [samples] to the ground truth dataset] (like a snowball) from the existing samples and it is used in situations where recruiting of samples is difficult.”)). “A POSITA would not have been motivated to combine a system suffering from a too-small sample size (*Deliu-Pendar*) with a system aimed at alleviating issues associated with ‘huge amount of training samples’ (*Xiao*).” EX2011 at ¶172 (citing *Xiao* at 335).

F. Ground 6: Obviousness in Light of *Deliu*, *Pendar*, and *Ukrainczyk*

Petitioner asserts that claims 8 and 9 are obvious in light of *Deliu*, *Pendar*, and *Ukrainczyk*.

1. Claim 8

Claim 8. The computer-implemented system of claim 1, wherein the data is parsed and segmented into tokens, and predetermined characters and predetermined common words are removed from the data.

Claim 8 depends on claim 1 and is not disclosed or suggested for the reasons set forth regarding claim 1.

Petitioner references its argument regarding claim element 1.a.4 for this claim. Pet. at 44. Patent Owner therefore incorporates its response thereto. This claim is not disclosed for the reasons set forth there.

2. Claim 9

Claim 9. The computer-implemented system of claim 1, wherein words within the data are reduced to their root form using stemming and lemmatization procedures.

Claim 9 depends on claim 1 and is not disclosed or suggested for the reasons set forth regarding claim 1.

Petitioner references its argument regarding claim 8 for this claim. Pet. at 59. Patent Owner therefore incorporates its response thereto. This claim is not disclosed for the reasons set forth there.

G. Ground 7: Obviousness in Light of *Deliu*, *Pendar*, and *Nunes*

Petitioner asserts that claim 11 is obvious in light of *Deliu*, *Pendar*, and *Nunes*.

1. Claim 11

Claim 11. The computer-implemented system of claim 1, further comprising: a crawler in operable communication with the processor and operable to traverse a plurality of deep web websites to retrieve Internet documents forming at least a portion of the deep web forum information.

Claim 11 depends on claim 1 and is not disclosed or suggested for the reasons set forth regarding claim 1.

Petitioner references its argument regarding claim 1 for this claim. Pet. at 61. Patent Owner therefore incorporates its response thereto. This claim is not disclosed for the reasons set forth there.

IX. SECONDARY CONSIDERATIONS OF NON-OBVIOUSNESS

Strong secondary considerations of non-obviousness support the '900 patent. These must be evaluated in assessing obviousness, to provide real-world insight into the novelty of the invention. *Graham v. John Deere Co. of Kan. City*, 383 U.S. 1, 17–18 (1966). These important factors counter hindsight bias and are particularly important in close cases. *Id.*

A. Commercial Success

The '900 patent has been commercialized by co-inventor Dr. Shakarian in his licensed company Cyber Reconnaissance, Inc. (“CYR3CON”) starting in 2016.

CYR3CON specifically implemented the patented features of the '900 patent in its PR1ORITY platform for predictive cyber-threat intelligence. Declaration of Dr. Shakarian (EX2002) at ¶7. Specifically, the product incorporated some of the key systems and methods claimed in the '900 patent for retrieving information from hacker forums on the dark web and deep web, enabling prioritization of patching to mitigate threats. *Id.* at ¶¶7 and 8 (“The ['900 patent] technologies were fundamental to the success of the CYR3CON PR1ORITY product”); *see also* EX2012 (CYR3CON’s PR1ORITY product roadmap) at 3, 11, 15, 16 (referencing hacker communications and multi-source data collection from the '900 patent). These technologies were implemented together in CYR3CON’s PR1ORITY commercial platform and its supporting backend, which delivered predictive cyber-threat intelligence to enterprise and OEM customers through multiple deployment channels. See EX2002 at ¶12.

CYR3CON’s PR1ORITY platform achieved meaningful commercial success directly attributable to the features claimed in the '900 patent and the related portfolio. EX2002 at ¶15. The platform integrated the '900 patent’s data collection and classification models to enable predictive models to assess exploitation likelihood based on hacker communications and social network metrics, providing unique proactive threat intelligence. EX2002 at ¶¶6, 7, and 11. Customers included

██████████ EX2002 at ¶13. The technology was also licensed, including to ██████████
██████████. *Id.* CYR3CON generated approximately ██████████
██████████
██████████ before its acquisition by Cyber Security Works (now Securin Inc.) in 2022.¹⁴ EX2002 at ¶14.

This success is directly tied to the patented technology: the '900 patent's fundamental enabling of multi-source dark web data collection and classification enabled high-accuracy predictions, which was essential for the platform's ability to identify emerging threats ahead of exploitation. See EX2002 at ¶15. The hierarchical tagging and automated classification of underground-forum content were critical to the system's ability to identify emerging threats and inform predictive cyber-risk assessments, and, to Dr. Shakarian's knowledge, no comparable commercial product offered equivalent functionality at the time. EX2002 at ¶15.

B. Long-Felt Need and Government Support

The '900 patent addressed a long-felt need for effective prediction of cyber exploits in a rapidly evolving threat landscape, critical to U.S. national security.

¹⁴ This paragraph contains confidential information that has been redacted pursuant to Patent Owner's pending motion to seal. *See* Paper 9.

EX2002 at ¶¶5-6. For that reason, Dr. Shakarian's research was funded by competitive government grants totaling over \$7 million, including from DARPA, IARPA, and ONR, focused on mission-critical cybersecurity. EX2002 at ¶5. This support highlights the invention's role in solving persistent challenges in threat prediction, where prior approaches lacked the integration of dark web data and machine learning for accurate prediction.

C. Failure Of Others

The prior art at the time failed to achieve the inventions of the '900 patent. Others explicitly discouraged various concepts claimed in the '900 patent. For example, *Deliu* thus teaches away from using a hierarchical classification scheme as in the '900 patent. EX2011 at ¶91. Similarly, *Frank* discourages and teaches away from the probability based indiscriminate child-node pruning of the '900 patent, and from using a single probability metric. *Id.* at ¶¶96 and 161. Thus, others attempted to solve the problem but failed.

D. Unexpected Results

Deliu and *Frank* teach away from various aspects of the invention, confirming that the results were surprising and not predicted.

E. Skepticism of Experts

The same teaching away confirms that leading experts in the field doubted the invention.

F. Praise By Others

The fact that Dr. Shakarian's research that led to the '900 patent was repeatedly funded by competitive government grants totaling over \$7 million, including from DARPA¹⁵, IARPA¹⁶, and ONR¹⁷, supports that industry experts acknowledged the significance of the invention. EX2002 at ¶4.

G. Copying By Competitors

Skysong's suits against CrowdStrike and Fortinet confirm that competitors have copied the invention, which suggests they recognized its value and that it was not obvious to them.

X. CONCLUSION

The Board should deny institution because CrowdStrike has failed to demonstrate a reasonable likelihood of prevailing as to any challenged claim.

Dated: December 1, 2025

Respectfully submitted,

/s/ Jay P. Kesan

Jay P. Kesan (Reg. No. 37,488)
KEY KESAN DALLMANN PLLC
1050 Connecticut Avenue, N.W. Suite
500 Washington, D.C. 20036

¹⁵ Defense Advanced Research Projects Agency, U.S. Department of Defense agency that funds national security technology.

¹⁶ Intelligence Advanced Research Projects Activity, R&D arm of U.S. Director of National Intelligence.

¹⁷ Office of Naval Research, U.S. Navy and Marine Corps' science and technology programs.

jay.kesan@kkd-law.com

Telephone: (202) 772-1100

Alfonso G. Chan

Andrew Cochran

Cahill Gordon & Reindel LLP

900 16th Street, N.W., Suite 500

Washington, D.C. 20006

Telephone: (202) 862-8942

Attorneys for Patent Owner

SKYSONG INNOVATIONS, LLC

CERTIFICATE OF SERVICE

Pursuant to 37 C.F.R. §§ 42.6(e)(4) and 42.25(b), the undersigned certifies that on 1 December 2025, a complete copy of Patent Owner's Preliminary Response was filed electronically through the Patent Trial and Appeal Board's PTABE2E System and provided, via electronic service, to the Petitioner by serving the correspondence address of record.

Dated: December 1, 2025

/s/ Jay P. Kesan

Jay P. Kesan

CERTIFICATE OF WORD COUNT

Pursuant to 37 C.F.R. § 42.24, the undersigned certifies that the foregoing Patent Owner's Preliminary Response contains less than 14,000 words excluding the table of contents, table of authorities, certificate of service, certificate of word count, and appendix of exhibits. Patent Owner has relied on the word count feature of the word processing software used to create this paper in making this certification.

Dated: December 1, 2025

/s/ Jay P. Kesan
Jay P. Kesan