Combating the hate speech in Thai textual memes

Lawankorn Mookdarsanit[1], Pakpoom Mookdarsanit[2]
[1]Faculty of Management Science, Chandrakasem Rajabhat University, Bangkok, Thailand
[2]Faculty of Science, Chandrakasem Rajabhat University, Bangkok, Thailand

Article Info

ABSTRACT

Thai textual memes have been popular in social media, as a form of image information summarization. Unfortunately, many memes contain some hateful content that easily causes the controversy in Thailand. For global protection, the *hateful memes challenge* is also provided by *Facebook AI* to enable researchers to compete their algorithms for combating the hate speech on memes as one of *NeurIPS'20* competitions. As well as in Thailand, this paper introduces the Thai textual meme detection as a new research problem in Thai natural language processing (Thai-NLP) that is the settlement of transmission linkage between scene text localization, Thai optical recognition (Thai-OCR) and language understanding. From the results, both regular and irregular text position can be localized by one-stage detection pipeline. More scene text can be augmented by different resolution and rotation. The accuracy of Thai-OCR using convolutional neural network (CNN) can be improved by recurrent neural network (RNN). Since misspelling Thai words are frequently used in social, this paper categorizes them as synonyms to train on multi-task pre-trained language model.

Corresponding Author:

Lawankorn Mookdarsanit
Faculty of Management Science
Chandrakasem Rajabhat University
39/1 Rachadapisek Road, Chan Kasem District, Chatuchak, Bangkok, 10900, Thailand
Email: lawankorn.s@chandra.ac.th

1. INTRODUCTION

Longer than 15 years of the social media's scale [1], confliction and violent in Thailand has been rapidly continuing as one of the main national problems [2], reported by *Ministry of Higher Education, Science, Research and Innovation, Thailand* (*MHESI*)'s minister. For the *MHESI*'s researching policy and support, the solution concerning this problem are still required [2] to improve the nation's unity. Many controversies are easily provoked [3-4] by sharing the hate speech on social media [4]. Under the umbrella of *United Nations (UN)* [5], hate speech can be both legal and illegal [5-6] that is any direct/indirect attacking characteristics [6], to negatively defame, insult, mock or scoff a person or a group of people. As a solution, the empowering social media in Thailand (e.g., *Facebook*, *Twitter*, *Instagram* and *LINE*) [7] is seen as the largest information centre that totally plays the main role for the relief of controversy. Coupled with the short communication on social media, the *image information categorization* [8] is popular for the well-classified knowledge, known as "infographic" [8-9]. Beyond the infographical representation, "meme" [10] - a form of *image information summarization* (with some *joking stories* or *explaining by laughing actions*) [10-11], is also a flavor of Thai social users [12-13] for the story telling that is short and clear, instead of reading the full article, e.g., "*Mag-Ina* meme (Thai: แม่ลูกไปไหน-กลับมาทำไม) [14]". According to the text rotationality, Thai texts in the memes as shown in Figure 1 can be located in both regular and irregular position. Against the meme objective [15], some Thai textual memes on *Facebook* are easily misleaded to

create the malicious content [16] called "hate speech" [17] that is the main cause of the serious confliction and violentmultipication in Thailand.Since the hate speech widely distributed on social media is a global problem to the entire nation, the *Hateful Memes Challenge2020* [17] is launched by *Facebook Artificial Intelligence (Facebook AI)* [18-19] as a track in *Conference on Neural Information Processing Systems (NeurIPS'20)* [20], to invite researchers around the world to compete their hate speech detection algorithmswith a $100K prize pool.To that end, the hate speech detection in textual meme is still a new problem in artificial intelligence (AI) [21] as multimodal (textual description and image) classification in English textual memes.

Figure 1. Thai textual meme examples, scene text rotationality: (a) Regular text position, (b) Irregular text position (Note that: the textual memes with "rudeness" or "social controversy"were censored)

Formerly, hate speech was only detected by textual information [22] using language understanding based on NLP [23-24]. The repeated hate speech on social media could be seen as a spam [25] or phishing link [26]. Since social media provided a rich source for mining negative/positive Thai trends in other forms (e.g., images, emoji symbols, GIFs, stickers) as well as textual information, there were many social apps to create textual memes. Textual meme was a popular media to share on social media, especially for representing hateful information. Hate speech detection could be seen as an extension of optical character recognition (OCR) application that required image to text conversion (img2txt).

Obviously, the hateful meme detection was the meeting of computer vision (CV) for image information and natural language processing (NLP) for textual information, respectively. As it related tothe concrete Thai controversy, most hateful memes in Thai social were already included some hateful/bully Thai texts within the images that could be seen as a Thai printed character recognition (or Thai optical character recognition: Thai-OCR) application [27-28]. Thai-OCR had been researching longer than 29 years [29]. During the historical AI for Thailand in 90's [30], not only Thai-OCR [31-33] but also Thai handwritten recognition [34-35] was one of the traditional open topics in Thai natural language processing (Thai-NLP) [36] that was seen as the computer applied to Thai language [37]. Although neural network based model demonstrated the high recognition rate in Thai-OCR [33, 38-40], it consumed huge much of large computational resource. Many works were proposed to avoid the neural network. Thai characters could be recognized by rough sets [41-42], numerical feature extraction [43-44]. Support vector machine (SVM) also provided a state-of-the-art recognition accuracy [45] for a large number of target classes. For the well-known competition, the local Thai-OCR challenge was hosted by *Thailand's National Electronics and Computer Technology Center (NECTEC)* that invited all Thai researchers or students to compete their algorithms shown on this contest, called "*Benchmark for Enhancing the Standard of Thai Language Processing (BEST)*" [46].

Since Thai writing variants by different writers that caused Thai handwritten recognition seemed to be more challenge than Thai-OCR [47-48], the later competitions shifted into Thai handwritten recognition since BEST 2014. Howbeit, Thai-OCR had its own challenge; such the scene text localization [49] and recognition [50] in the wild that might have multi-objects within the scene as in the multi-views of Thai license plate recognition [51-52]. According to the scene text rotationality, Thai texts on the scene [53] could be detected [54-55] in both regular and irregular position (categorized by *Clova AI, LINE Corporation* [50]).

To set the transmission linkage between scene text localization [49, 51-55] with Thai-OCR [27-28, 31-33, 38-46] and Thai hateful text understanding [56-57] as a new Thai-NLP problem in Figure 2, this paper proposes an end-to-end "*combating the hate speech in Thai textual memes*" as a solution for the problem stated by *MHESI* and *Facebook AI*. As well as visual question answering (VQA) [58-59] and image captioning (IC) [60-61], this new problem needs both CV and NLP task. Since multi-objects can be located with Thai-texts in the same scene, the positions of texts are localized by single shot detector (SSD) [62]. For img2txt conversion, there are so many vertical positions (top, upper, middle and lower level) as sequence data [63] in Thai text that is neccasary for character-level embedding, coupled with character recognition [64] based on residual network (ResNet) [65] and bidirectional long-short-term-memory (Bi-LSTM) [66] with the connectionist temporal classification (CTC) [67]. The hate speech in a converted Thai text is finally detected by a Multi-tasking transfer learning architecture as generative pre-training transformer v.2 (GPT-2) [68] by OpenAI. The main contribution is made as follow:

a) Thai hateful meme detetection is proposed as the meeting between CV and NLP that poses a new research problem in Thai-NLP.
b) The regular Thai texts are multiplied by different power law distribution; and rotated in different angles to enlarge the dataset for training the irregular ones. Both regular and irregular Thai texts in memes can be detected by SSD.
c) The accuracy of Thai-OCR by ResNet in sequence data in character level can be improved by Bi-LSTM.
d) Those frequent misspelling words are seen as Thai synonyms that are combined with multi-task GPT-2 to produce the state-of-the-art results.

This paper is organized into 4 sections. Research method is described in Section 2. Section 3 is results and discussion. And the conclusion is in Section 4.
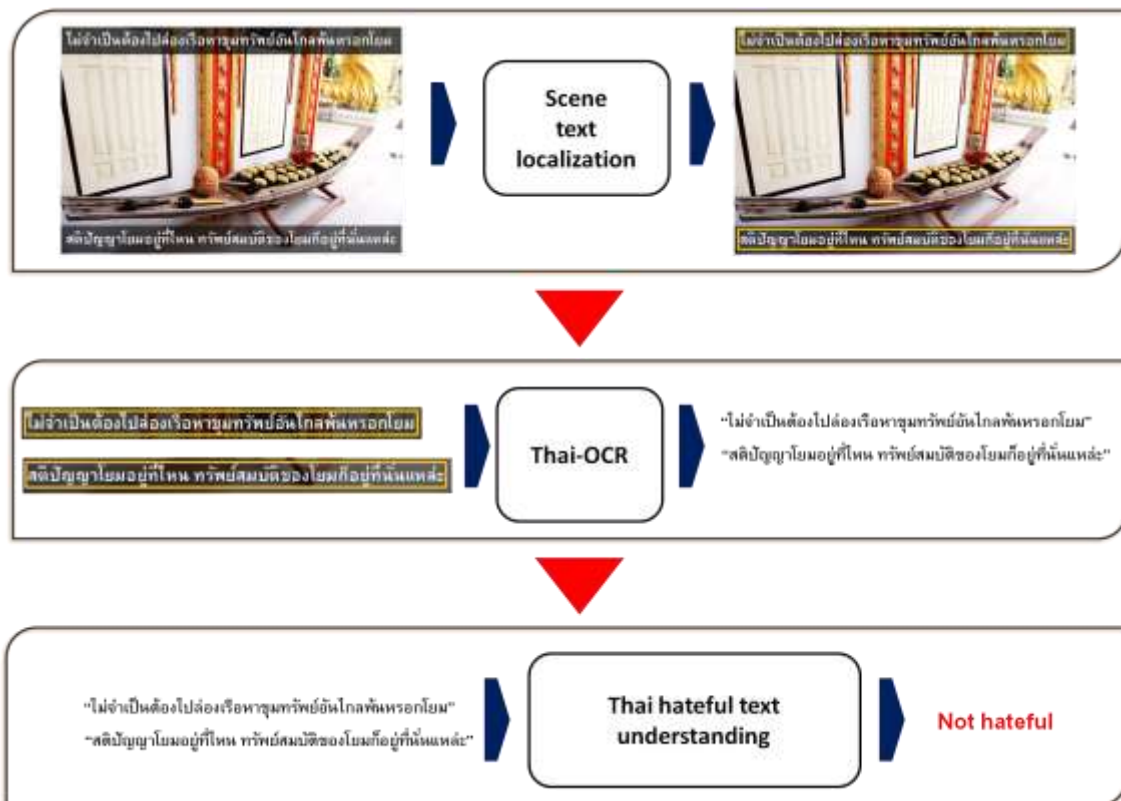

Figure 2. This proposed "*hate speech detection in Thai memes*" as the transmission linkage between scene text localization with Thai-OCR and Thai hateful text understanding

2. RESEARCH METHOD

Most story telling in hateful memes are normally illustrated by visual image features with the bully texts for satirizing or mocking the activities. To formulate Thai hate speech detection on memes as Figure 2, this paper poses an enlargement of Thai natural language processing (Thai-NLP). The modular framework consists of scene text localization-to localize the position of Thai texts within the meme, Thai-OCR-to convert Thai textual image into sequential characters of text incharacter-level (also called Thai-img2txt) and Thai hateful text understanding-to supervisedly learn and classify the converted Thai textual information using word-level sequence to sequence (seq2seq) network.

2.1. Scene text localization

For scene textual meme (unlike Thai-OCR problem), Thai characters cannot be directly segmented from the background. Thai text can be located in any positions within the meme. Moreover, there are a large number of objects with textual description in the scene (called scene complexity). Convolutional neural network (CNN) based detection has been proven to be better than traditional detection to localize Thai text [53] from the multi-objects scene. Revolutionarily, CNN-based detection can be classified into 2 types [69]: 1) *one-stage pipeline* (without region feature extraction) and *two-stage pipeline* (with region feature extraction). From the report [69-70], one-stage pipeline is better than two-stage in term of speed; but it provides lower in correctness. Two-stage detection is proposed to localize multi-objects, rather than text. Generally, Thai textual features can besufficiently localized by one-stage pipeline, according to the short time processing.

Single shot detector (SSD) [62] is such a one-stage pipeline that is shown to be the acceptable correctness of Thai text localization [49] in the multi-objects scenes. SSD is based on Visual geometry group network (VGGNet) with ImageNet pre-training [71]. SSD firstly provides the multi-reference and multi-resolution (called pyramid representation) in one-stage for various sizes of objects, especially for Thai texts. Accordingly, this paper uses SSD to localize the positions of Thai textual features from memes as text localization in the wild scene.

2.2. Thai-OCR

Thai optical character recognition (Thai-OCR) is based on convolutional neural network (CNN) for feature extraction; and recurrent neural network (RNN) for character-level sequence [64]. The localized Thai textual feature is processed in 2 stages: *Thai textual feature extraction* is used to recognize those Thai characters (44 constants, 18 vowels, 4 tone marks, 5 diacritics, 19 numbers and 6 symbols) [37, 46]. For Thai character recognition, the block is automatically divided into 4 levels (top, upper, medium and lower position) as shown in Figure 3(a). The residual network (ResNet) [65] pretrained on ImageNet is applied to supervisedly-learn and classify those characters from the localized textual features.

Thai character-level sequence is to sequence the extracted textual features by Bidirectional Long-short-term memory (Bi-LSTM) [66].Each character feature is sequentially sorted from left to right as position by position. Each position is also checked its lower, upper and top level, as shown in Figure 3(b). The stream extracted features is fed into Bi-LSTM [66]. For the prediction, Connectionist temporal classification (CTC) [67] is used for mapping those extracted features into Thai character sequence, to produce the output as Thai textual information.
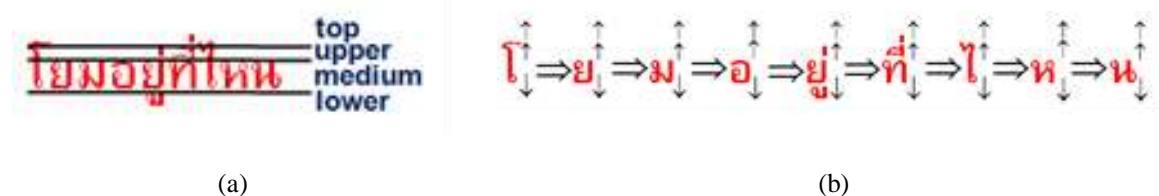

(a) (b)

Figure 3. Thai character-level sequence: (a) Thai characters in 4 levels: top, upper, medium and lower position, (b) Sorting each character feature from left to right; and each of the lower, upper and top level

2.3. Thai hateful text understanding

According to the infographic representation, most Thai textual memes are short description. Speech understanding can be seen as a problem of aspect-level sentiment analysis [72] on Thai textual information that consists of *frequent misspelling words* and *pre-trained language model*.

2.3.1. Frequent misspelling words

Although Thai textual information in meme is such a short textual message, the real Thai usage still has a huge much of noise and character repetition that are the main cause of model misclassification. Those spamming characters are filtered out using pyThaiNLP library [73], except for frequent misspelling words. Likewise, Thai word tokenization is still an important issue in Thai natural language processing (Thai-NLP) for the word-level language understanding. Instead of the misspelling word checker, the frequent misspelling usages can be seen as synonyms on social media, as shown in Table 1. These frequent misspelling usages are also augmented to be trained to the multi-task fine-tuning in pre-trained language model.

Table 1. Some official spelling and frequent misspelling

Official Thai spelling	Frequent misspelling forms
บอกตรงๆ	บ่องตง
ครับ	คร้าฟ, ครัช, ครัฟ, คับ, คัฟ, ฮ้าฟ, งับ, ฮัฟ
ค่ะ/คะ	ค่า, ค้า, ขา, ค๊ะ, ข๊ะ, ข่ะ
รำคาญ	รามคาญ, รัมคาน, รำค่าน, รามคาน
อะไร	อะรัย, อาไร, อาราย
ทำไม	ทามมาย, ทัมมัย, ทำมัย, ทะมาย
จังเลย	จงเบย, จังเบย, จังเรย, จางเรย
อะไร	อะรัย, อัลไล, อาราย, อาไร

2.3.2. Pre-trained language model

Pre-trained language model has demonstrated state-of-the-art results [74] in word2vec language understanding tasks for Thai text classification [75]. Generative pre-training transformer 2 (GPT-2) by OpenAI is used to be pre-trained and fine-tuned Thai textual information from the collected dataset. For the benchmarking, Wisesight sentiment analaysis [76] in Kaggle competition [77] that has 4 different target classes: positive, negative, neutral and question. GPT-2 is an improved version of GPT [78] for semi-supervised learning on large-scale dataset (both labeled and unlabled data) that has transformer based model and deeper than GPT. GPT-2 also has multi-tasked fine-tuning at the same time to improve the performance of standard transfer learning. The fine-tuned configuration is set for 3 epochs and batch size as 32 with 1.5B parameters. To best of the knowledge, the frequent misspelling Thai words are augmented to fine-tune the model as multi-tasking.

3.   RESULTS AND DISCUSSION

According to the requirement of Thai hateful meme detection on social media (as well as the hateful memes challenge by *Facebook AI*), this section described the implementation detail that processed on Tesla V100 GPU colab. By the components, the results could be separately divided into *dataset and extension*, *Thai textual meme results* and *language understanding results*.

3.1.   Dataset and extension

The dataset in this paper could be divided into 2 parts: *textual image dataset* and *textual information dataset*. Both dataset had the augmentation to increase more synthetic data.

3.1.1. Textual image dataset

A various amount of cleaned Thai textual images (without any other objects) in different fonts, e.g., dhamma, religion, politics, celebrities, business, science, quotes or hate speech were cropped. To increase the dataset, these textual formats were multiplied by power law distribution (where $0 \leq \deg ree \leq 1$)to increase more multi-resolution texts and rotated in both left and right direction (where $30° \leq \theta_{\Leftrightarrow} \leq 60°$) to extensively synthesize more irregular texts, according to the textual presentation for the reader's views, as shown in Figure 4. The textual image dataset contained 16,469 images with their textual information as the target class.

(a)   (b)

Figure 4. Synthesizing more textual images: (a) Multi-resolution degree, (b) Rotation

3.1.2. Textual information dataset

For pre-training dataset construction, the 325,530 raw Thai texts concerning opinion analysis were crawled from *Pantip.com*, *Youtube*, *Facebook* and *Twitter* and were cleaned and prepared that finally had only 298,212 texts for the pre-trained model. Instead of misspelling replacement, the frequent misspelling words were seen as synonyms; those synonyms might be further used to extensively synthesize another 72,398 Thai texts for multi-task fine-tuning. By the way, Wisesight sentiment analysis [76-77] in Kaggle competition [77] was used as benchmarking dataset for the proposed language model that had 26,737 Thai texts.

3.2. Thai textual meme results

The training set and testing set was partitioned into 70:30. According to object detection and classification in computer vision, Thai optical character recognition (Thai-OCR, called img2txt conversion) results could be divided into *text localization* and *character recognition evaluation*.

3.2.1. Text localization evaluation

Thai textual with multi-objects in the scene (called scene complexity) that made the text localization in this problem differ from classical Thai-OCR. And those Thai characters (44 constants, 18 vowels, 4 tone marks, 5 diacritics, 19 numbers and 6 symbols) were not such a complex feature that was unnecessary to use two-stage detection, e.g., Faster R-CNN [79], FPN [80]. As to the short time localization, one-stage detection (e.g., YOLO [81], SSD [62]) was more suitable for these characters with the careless region feature formulation. From the reported results [70], SSD was the best speed in object detection; one-stage detection was quicklier than two-stage detection as shown in Figure 5. Text localization correctness by SSD [62] could be divided into regular, irregular and combined results as shown in Table 2. Since many Thai textual memes were little orientated the text position, it was essential to extend the efficiency by synthesizing more Thai textual samples to the deep learning model.
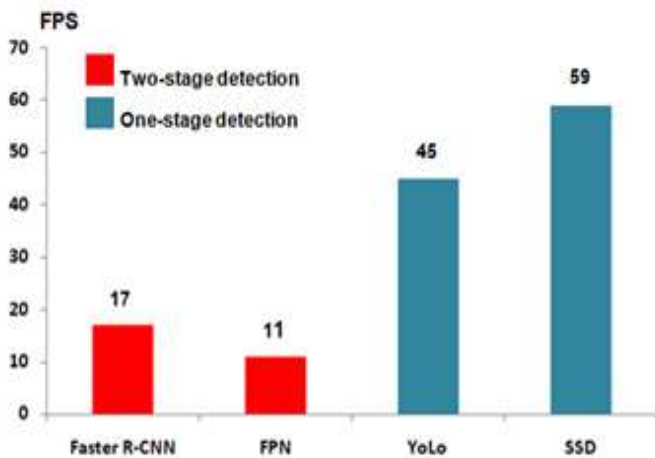
Table 2. Comparison between regular and irregular scene text detection

Thai text position	Precision
Regular	93.2
Irregular	88.4

Figure 5. Reported speed (in FPS) comparison between SSD and other detection pipelines

3.2.2. Text recognition evaluation

All Thai character features were trained and recognized by ResNet architecture [65] and CTC prediction [67]. From the results, Bi-LSTM [66] totally improved the recognition accuracy in character-level sequence data, even if it took almost twice times. Many text recognition papers used ResNet or VGGNet. With a larger number of parameters and skip connection in ResNet, ResNet outperformed VGGNet. For trade-off between correctness and speed, the higher recognition accuracy (%) was paid by more processing time (ms/image). Accuracy and time curve were shown in Figure 6. While, the overall character-level recognition results under CTC prediction were shown in Table 3.
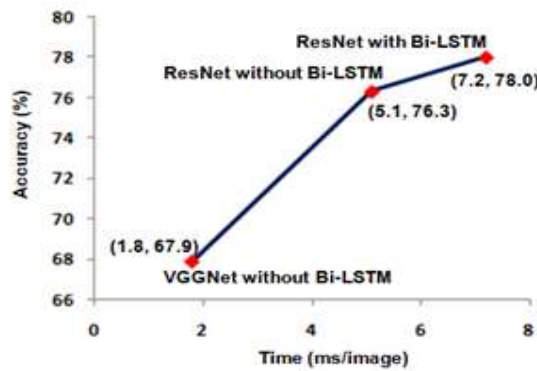

Figure 6. Accuracy and time curve

Table 3. The overall character-level recognition results under CTC prediction

Thai textual feature extraction	Character-level sequence	Time	Accuracy	Parameters
VGGNet	None	1.8	67.9	5.6M
ResNet	None	5.1	76.3	46M
ResNet	Bi-LSTM	7.2	78.0	48M

3.3. Thai textual understanding results

The data partitioning in pre-training was divided into 3 parts: training, validation and testing dataset as 70:15:15. For fine-tuning, the training and testing dataset was splitted as 70:30. From Table 4, standard GPT-2 might look better than GPT-2 with multi-tasking in pre-training stage. In contrast, GPT-2 with multi-tasking was fine-tuned in multi-task fine-tuning those spelling and frequent misspelling words as the different written styles to be a better final accuracy in Wisesight's private and public dataset as 0.7634 and 0.7832.

Multi-tasking or (multi-task fine-tuning at the same time) totally boosted the fairness in augmented transformer-based model that no tasks had higher precedence than others. Based on Wisesight challenge, the promising results were demonstrated that multi-task based transfer learning improved the classification accuracy in Thai textual categorization.

Table 4. Comparisson between word-level GPT-2 with/without multi-taksing

Language model	Wisessight Challenge	
	Private	Public
Multi-tasking GPT-2	0.7634	0.7832
GPT-2	0.6891	0.7011

4. CONCLUSION

As referred to *Ministry of Higher Education, Science, Research and Innovation, Thailand (MHESI)*'s requirement on social media literacy, this paper achieved combating the hate speech in Thai textual memes as a new research problem in Thai natural language processing (Thai-NLP). The research method contained 3 parts: scene text localization by single shot detector (SSD), Thai optical character recognition (Thai-OCR) by residual network (ResNet) and Bidirectional Long-short-term-memory (Bi-LSTM) and Thai hateful text understanding by multi-task GPT-2. For the main discovery, the regular text multi-resolution degree and rotation for affine augmentation could improve the localization and recognition.

Bi-LSTM improved the recognition accuracy in character-level sequence data. Instead of cleansing Thai word misspelling, turn them as synonyms to multi-task GPT-2. According to the nature of Thai comparison types, many sarcastical objects within the image (e.g., criminals, buffalos, dogs, garbages or other caricatures) are also useful for combating the hateful memes. Moreover, the Deepfake can dangerously generate the multi-emotions from a facial person that will be easily applied for making the hateful meme. The fake face detection will be convergent to hateful meme detection as the same research goal.

ACKNOWLEDGEMENTS

REFERENCES

[1] D. Hendricks, "Complete History of Social Media: Then and Now," 2013. [Online]. Available: https://smallbiztrends.com/2013/05/the-complete-history-of-social-media-infographic.html. Accessed: September 24, 2020.

[2] MHESI. "Media's Rights and Liberties Are Necessary," 2020. [Online]. Available: https://www.mhesi.go.th/index.php/en/all-media/infographic/2757-2020-11-22-07-44-13.html. Accessed: September 24, 2020. [in Thai].

[3] W. Schaffar, "New Social Media and Politics in Thailand: The Emergence of Fascist Vigilante Groups on Facebook," *Austrian Journal of South-East Asian Studies*, vol. 9, no. 2, pp. 215-233, 2016.

[4] Xinhua News Agency, "Gov't Warns Social Media, Websites over Publishing 'MisInfo," Khaosod English, 2020. [Online]. Available: https://www.khaosodenglish.com/news/crimecourtscalamity/2020/08/21/govt-warns-social-media-websites-over-publishing-misinfo/. Accessed: September 24, 2020.

[5] United Nations, "United Nations Strategy and Plan of Action on Hate Speech," pp. 1-5, 2019. [Online]. Available: https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf. Accessed: September 24, 2020.

[6] C. George, "Hate Speech Law and Policy," *The International Encyclopedia of Digital Communication and Society*, First edition, John Wiley & Sons, Inc, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118767771.wbiedcs139.

[7] N. Mahittivanicha, "Most Active Social Media Platforms in Thailand during July 2020," 2020. [Online]. Available: https://www.twfdigital.com/blog/2020/04/summary-of-social-network-in-thailand-march2020/. Accessed: September 24, 2020. [in Thai].

[8] J. Lankow, et al., "Infographics: The Power of Visual Storytelling,"John Wiley & Sons, Inc, 2012.

[9] M. K. Mancini, "Infographic: Social Media, Then and Now," 2015. [Online]. Available: https://www.tech4pub.com/2015/07/08/infographic-social-media-then-and-now/. Accessed: September 24, 2020.

[10] C. M. CastañoDíaz, "Defining and Characterizing the Concept of Internet Meme," *Revista CES Psicología*, vol. 6, no. 1, pp. 82-104, 2013.

[11] B. H. Spitzberg, "Toward a Model of Meme Diffusion (M3D)," *Communication Theory*, vol. 24, no. 3, pp. 311-339, 2014.

[12] V. Taecharungroj and P. Nueangjamnong, "The Effect of Humour on Virality: The Study of Internet Memes on Social Media," in *the 2014 International Forum on Public Relations and Advertising Media Impacts on Culture and Social Communication*, 2014.

[13] V. Taecharungroj and P. Nueangjamnong, "Humour 2.0: Styles and Types of Humour and Virality of Memes on Facebook," *Journal of Creative Communications*, vol. 10, no. 3, pp. 288-302, 2015.

[14] Thaitrakulpanich, A., "Filipino Mom-and-kid 'Mag-Ina' Meme Spreads to ThaiNet," Khaosod English, 2019. [Online]. Available: https://www.khaosodenglish.com/culture/net/2019/12/12/filipino-mom-and-kid-mag-ina-meme-spreads-to-thainet/. Accessed: September 24, 2020.

[15] L. Shifman, "Memes in a Digital World: Reconciling with a Conceptual Troublemaker," *Journal of Computer-Mediated Communication*, vol. 18, pp. 362-377, 2013.

[16] R. Sittichai and P. K. Smith, "Bullying and Cyberbullying in Thailand: Coping Strategies and Relation to Age, Gender, Religion and Victim Status," *Journal of New Approaches in Educational Research*, vol. 7, no. 1, pp. 24-30, 2018.

[17] D. Kiela, et al., "The Hateful Memes Challenge: Detecting Hate Speech in Multimodal Memes," 2020. [Online]. Available: https://arxiv.org/pdf/2005.04790.pdf. Accessed: September 24, 2020.

[18] Facebook AI. "Hateful Memes Challenge and Data Set," 2020. [Online]. Available: https://ai.facebook.com/tools/hatefulmemes/. Accessed: October 4, 2020.

[19] Facebook AI. "Hateful Memes Challenge and Data Set for Research on Harmful Multimodal Content," 2020. [Online]. Available: https://ai.facebook.com/blog/hateful-memes-challenge-and-data-set/. Accessed: October 4, 2020.

[20] NeurIPS. "NeurIPS 2020 Competition Track," 2020. [Online]. Available: https://neurips.cc/Conferences/2020/CompetitionTrack. Accessed: October 4, 2020.

[21] H. Zhong, et al., "Content-Driven Detection of Cyberbullying on the Instagram Social Network," in *the 2016 International Joint Conference on Artificial Intelligence*, pp. 3952-3958, 2016.

[22] Z. Waseem, et al., "Understanding Abuse: A Typology of Abusive Language Detection Subtasks," in the *2017 Workshop on Abusive Language Online*, pp. 78-84, 2017.

[23] P. Fortuna and S. Nunes, "A Survey on Automatic Detection of Hate Speech in Text," *ACM Computing Surveys*, vol. 51, no. 4, 2018.

[24] A. Schmidt and M. Wiegand, "A Survey on Hate Speech Detection using Natural Language Processing," in the *2017 International Workshop on Natural Language Processing for Social Media*, pp. 1-10, 2017.

[25] Y. Vernanda, et al., "Indonesian Language Email Spam Detection using N-gram and Naïve Bayes Algorithm," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 5, pp. 96-108, pp. 2012-2019, 2020.

[26] J. A. Jupin, et al., "Review of the Machine Learning Methods in the Classification of Phishing Attack," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 4, pp. 1545-1555, 2019.

[27] T. Siriteerakul, et al., "Character Classification Framework based on Support Vector Machine and k-Nearest Neighbour Schemes," *Science Asia*, vol. 42, no. 1, pp. 46-51, 2016.

[28] K. Kesorn, et al., "Optical Character Recognition (OCR) Enhancement using an Approximate String Matching Technique," *Engineering and Applied Science Research*, vol. 45, no. 4, pp. 282-289, 2018.

[29] V. Sornlertlamvanich, "A 29-year Journey of Thai NLP," [Online].Available: https://www.slideshare.net/virach/a-29year-journey-of-thai-nlp. Accessed: October 4, 2020.

[30] B. Kijsirikul and T. Theeramunkong, "Survey on Artificial Intelligence Technology in Thailand," 1999. [Online]. Available: https://www.cp.eng.chula.ac.th/~boonserm/publication/AISurvey.pdf. Accessed: October 4, 2020.

[31] C. Kimpan and S. Walairacht, "Thai Characters Recognition," in the *1994 Symposium on Natural Language Processing*, 1994.

[32] D. Cooper, "How Do Thais Tell Letters Apart?," *1996 International Symposium on Language and Linguistics*, 1996.

[33] C. Tanprasert and T. Koanantakool, "Thai OCR: a Neural Network Application," in the *1996 Digital Processing Applications*, pp. 90-95, 1996.

[34] C. Lursinsap and C. Khunasaraphan, "Simulated Light Sensitive Model for Handwritten Digit Recognition," in *IJCNN Joint Conference on Neural Networks*, pp. 13-18, 1992.

[35] C. Khunasaraphan and C. Lursinsap, "Simulated Light Sensitive Model for Thai Handwritten Alphabets Recognition," in the *1994 Symposium on Natural Language Processing*, 1994.

[36] V. Sornlertlamvanich, et al., "The State of the Art in Thai Language Processing," in the *2000 Annual Meeting of Association for Computational Linguistics*, pp. 1-2, 2000.

[37] H. T. Koanatakool, et al., "Computers and the Thai Language," *IEEE Annals of the History of Computing*, vol. 31, no. 2, pp. 50-58, 2009.

[38] C. Tanpraset, et al., "Improved Mixed Thai & English OCR using Two-step Neural Net Classification," *NECTEC Technocal Journal*, vol. 1, no. 1, pp.41-46, 1996.

[39] B. Kijsirikul, et al., "Thai Printed Character Recognition by Combining Inductive Logic Programming with Backpropagation Neural Network," in *IEEE Asia-Pacific Conference on Circuits and Systems Microelectronics and Integrating Systems*, 1998.

[40] U. Marang, et al., "Recognition of Printed Thai Characters using Boundary Normalization and Fuzzy Neural Networks," in the *2000 Symposium on Natural Language Processing*, 2000.

[41] W. Kasemsiri and C. Kimpan, "Printed Thai Character Recognition using Fuzzy-Rough Sets," in *IEEE Region 10 International Conference on Electrical and ElectronicTechnology*, 2001.

[42] P. Phokharatkul and P. Pantaragphong, "Printed Thai Characters Recognition using Rough Sets and Fractal Dimension," in *International Symposium on Nonlinear Theory and Its Applications*, 2002.

[43] A. Kawtrakul and P. Waewsawangwong, "Multi-feature Extraction for Printed Thai Character Recognition," in the *2000 Symposium on Natural Language Processing*, 2000.

[44] U. Suttapakti, et al., "Font Descriptor Construction for Printed Thai Character Recognition," in the *2013 IAPR International Conference onf Machine Vision Applications*, 2013.

[45] T. Siriteerakul, et al., "Character Classification Framework based on Support Vector Machine and k-Nearest Neighbour Schemes," *Science Asia*, vol. 42, no. 1, pp. 46-51, 2016.

[46] I. Methasate and S. Marukatut, "BEST 2013: Thai Printed Character Recognition Competition," in the *2013 Symposium on Natural Language Processing*, 2013.

[47] O. Surinta, et al., "Recognition of Handwritten Characters using Local Gradient Feature Descriptors," *Engineering Applications of Artificial Intelligence*, vol. 15, no. 1, pp. 405-414, 2015.

[48] P. Mookdarsanit and L. Mookdarsanit, "ThaiWrittenNet: Thai Handwritten Script Recognition using Deep Neural Networks," *Azerbaijan Journal of High Performance Computing*, vol. 3, no. 1, pp. 75-93, 2020.

[49] Y. Baek, et al., "Character Region Awareness for Text Detection," in the *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9357-9366, 2019.

[50] J. Baek, et al., "What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis," in the *2019 IEEE/CVF Conference on Computer Vision*, pp. 4714-4722, 2019.

[51] W. Puarungroj and N. Boonsirisumpun, "Thai License Plate Recognition based on Deep Learning," in the *2018 International Conference on Computer Science and Computational Intelligence*, pp. 214-221, 2018.

[52] A. Kitvimonrat and S. Watcharabutsarakham, "A Robust Method for Thai License Plate Recognition," in the *2020 International Conference on Image and Graphics Processing*, pp. 17-21, 2020.

[53] T. Kobchaisawat and T. H. Chalidabhongse, "Thai Text Localization in Natural Scene Images using Convolutional Neural Network," in the *2014 Signal and Information Processing Association Annual Summit and Conference*, pp. 1-7, 2014.

[54] X. Zhou, et al., "EAST: An Efficient and Accurate Scene Text Detector," in the *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5551-5560, 2017.

[55] T. Kobchaisawat, et al., "Scene Text Detection with Polygon Offsetting and Border Augmentation," *Electronics*, vol. 9, no. 117, 2020.

[56] S. Sirihattasak, et al., "Annotation and Classification of Toxicity for Thai Twitter," in the *2018 Text Analytics for Cybersecurity and Online Safety*, 2018.

[57] P. Mookdarsanit and L. Mookdarsanit, "TGF-GRU: A Cyber-bullying Autonomous Detector of Lexical Thai across Social Media," *NKRAFA Journal of Science and Technology*, vol. 15, no. 1, pp. 50-58, 2019.

[58] M. Malinowski, et al., "Ask Your Neurons: A Neural-Based Approach to Answering Questions about Images," in the *2015 IEEE International Conference on Computer Vision*, pp. 1-9, 2015.

[59] K. Marino, et al., "OK-VQA: A Visual Question Answering Benchmark Requiring External Knowledge," in the *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3190-3199, 2019.

[60] K. Shuster, et al., "Engaging Image Captioning via Personality," in the *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12508-12518, 2019.

[61] P. Mookdarsanit and L. Mookdarsanit, "Thai-IC: Thai Image Captioning based on CNN-RNN Architecture," *International Journal of Applied Computer and Information Systems*, vol. 10, no. 1, pp. 40-45, 2020.

[62] W. Liu, et al., "SSD: Single Shot MultiBox Detector," in the *2016 European Conference on Computer Vision*, pp. 21-37, 2016.

[63] G. W. Nie, et al., "Deep Stair Walking Detection using Wearable Inertial Sensor via Long Short-Term Memory Network," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 1, pp. 238-246, 2020.

[64] T. Emsawas and B. Kijsirikul, "Thai Printed Character Recognition using Long Short-Term Memory and Vertical Component Shifting," in the *2016 Pacific Rim International Conference on Artificial Intelligence*, pp. 106-115, 2016.

[65] K. He, et al., "Deep Residual Learning for Image Recognition," in the *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.

[66] A. Graves, et al., "Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition," in the *2005 International Conference on Artificial Neural Networks*, pp. 799-804, 2005.

[67] A. Graves, et al., "Connectionist Temporal Classification: Labeling Unsegemented Sequence Data with Recurrent Nerual Networks," in the *2006 International Conference on Machine Learning*, pp. 369-376, 2006.

[68] A. Radford, et al., "Language Models are Unsupervised Multitask Learners," 2019. [Online]. Available: https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf. Accessed: October 4, 2020.

[69] L. Liu, et al., "Deep Learning for Generic Object Detection: A Survey," 2018. [Online]. Available: https://arxiv.org/pdf/1809.02165.pdf. Accessed: October 4, 2020.

[70] Z. Zou, et al., "Object Detection in 20 Years: A Survey," 2019. [Online]. Available: https://arxiv.org/pdf/1905.05055.pdf. Accessed: October 4, 2020.

[71] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-scale Image Recognition," in the *2015 International Conference on Learning Representations*, 2015.

[72] A. Torfi, et al., "Natural Language Processing Advancements by Deep Learning: A Survey," 2020. [Online]. Available: https://arxiv.org/pdf/2003.01200.pdf. Accessed: October 4, 2020.

[73] PyThaiNLP. "PyThaiNLP: Thai Natural Language Processing in Python," 2019. [Online]. Available: https://github.com/PyThaiNLP/pythainlp. Accessed: October 4, 2020.

[74] M. Raghu and E. Schmidt, "A Survey of Deep Learning for Scientific Discovery," 2020. [Online]. Available: https://arxiv.org/pdf/2003.11755.pdf. Accessed: October 4, 2020.

[75] T. Horsuwan, et al., "A Comparative Study of Pretrained Language Models on Thai Social Text Categorization," in the *2020 Asian Conference on Intelligent Information and Database Systems*, pp. 63-75, 2020.

[76] PyThaiNLP, "wisesight-sentiment," 2018. [Online]. Available: https://github.com/PyThaiNLP/wisesight-sentiment. Accessed: October 4, 2020.

[77] Kaggle, "Wisesight Sentiment Analysis," 2018. [Online]. Available: https://www.kaggle.com/c/wisesight-sentiment. Accessed: October 4, 2020.

[78] A. Radford, et al., "Improving Language Understanding by Generative Pre-Training," 2018. [Online]. Available: https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf. Accessed: October 4, 2020.

[79] S. Ren, et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.doi: 10.1109/TPAMI.2016.2577031.

[80] T-Y. Lin, et al., "Feature Pyramid Networks for Object Detection," in the *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 936-944, 2017.doi: 10.1109/CVPR.2017.106.

[81] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in the *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6517-6525, 2017.doi: 10.1109/CVPR.2017.690.