



A Systematic Literature Review Using Artificial Intelligence and Fake News Detection

Sarawut Kedtarwon

Valaya Alongkorn Rajabhat University under the Royal Patronage Pathum Thani Province, Thailand

E-mail: sarawut.ked@vru.ac.th, ORCID ID: <https://orcid.org/0009-0009-8583-1973>

Received 25/11/2023

Revised 02/12/2023

Accepted 10/12/2023

Abstract

Background and Aims: The Thai population uses the internet in large numbers as a medium for receiving news. Fake news is distributed on social media in large numbers. Artificial intelligence was born. Computers can learn on their own and be able to infer what humans have created. is natural language processing Analyzing various data and creating models from that data using machine learning. The purpose of this article is 1) To study methods and models for detecting fake news using artificial intelligence. 2)To propose guidelines for using artificial intelligence to detect fake news.

Methodology: The literature review is Transparent and by scientific principles, which can be proven repeatedly, the researchers therefore adopted a process of systematically reviewing the literature. Applied to this research is divided into 3 steps: 1) define the search issues, 2) review the literature and check the quality of the research, and 3) synthesize the information to summarize into a body of knowledge. The criteria for considering literature will be considered sequentially, divided into 4 sequences: 1) Research published within the last 5 years 2) Considering the names of literature, coming directly to the topic of artificial intelligence fake news. and natural language processing 3) Literature is in English format 4) Consider literary abstracts The criteria for consideration include detecting fake news using artificial intelligence processes and considering the full literature Therefore, a study was conducted from research on detecting fake news with artificial intelligence, a total of 10 cases.

Results: The artificial intelligence can detect fake news correctly, it is highly satisfactory at more than 90 percent and has guidelines for using artificial intelligence to detect fake news. However, the researcher has an observation. Data analysis used in computer teaching will always fit that model, but on social media, there is news happening every second. Therefore, it is necessary to study and develop artificial intelligence to detect fake news. More to detect fake news with up-to-date artificial intelligence.





Conclusion: Artificial intelligence can detect fake news with an accuracy of more than 90%, and there are guidelines for using artificial intelligence to detect fake news. However, the use of data in teaching computers must be based on the nature of the model. And because news on social media happens every second, further development of artificial intelligence to detect fake news in a timely manner is necessary.

Keywords: Artificial Intelligence; Natural Language Processing; Fake News



การทบทวนวรรณกรรมอย่างเป็นระบบโดยใช้ปัญญาประดิษฐ์กับการตรวจสอบข่าวปลอม

ศราวดี เกิดถาวร

มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์จังหวัดปทุมธานี

บทคัดย่อ

ภูมิหลังและวัตถุประสงค์: ประชากรไทยมีการใช้อินเทอร์เน็ตเป็นจำนวนมากในการเป็นสื่อกลางในการรับข่าวสาร ข่าวปลอม กระจายในสื่อสังคมออนไลน์เป็นจำนวนมาก ปัญญาประดิษฐ์ถือกำเนิดขึ้นเครื่องคอมพิวเตอร์มีความสามารถที่จะเรียนรู้เองสามารถสรุปสิ่งที่มนุษย์สร้างขึ้นได้ คือการประมวลผลภาษาธรรมชาติ การวิเคราะห์ข้อมูลต่าง ๆ และการสร้างแบบจำลองจากข้อมูลเหล่านั้นด้วยศาสตร์ทางด้านการเรียนรู้ของเครื่อง วัตถุประสงค์ของบทความนี้คือ 1) เพื่อศึกษาวิธีการและแบบจำลองการตรวจสอบข่าวปลอมด้วยปัญญาประดิษฐ์ 2) เพื่อเสนอแนวทางการใช้ปัญญาประดิษฐ์ตรวจสอบข่าวปลอม เพื่อให้การทบทวนวรรณกรรมโปร่งใสและตรงตามหลักการทางวิทยาศาสตร์ซึ่งสามารถพิสูจน์ซ้ำได้

ระเบียบวิธีการวิจัย: เป็นการนำกระบวนการในการทบทวนวรรณกรรมอย่างเป็นระบบมาประยุกต์ใช้กับงานวิจัยนี้ โดยแบ่งออกเป็น 3 ขั้นตอน 1) กำหนดประเด็นการสืบค้น 2) ทบทวนวรรณกรรมและตรวจสอบคุณภาพงานวิจัย 3) สังเคราะห์ข้อมูลเพื่อสรุปเป็นองค์ความรู้ ซึ่งเกณฑ์ในการพิจารณาวรรณกรรมจะถูกพิจารณาเป็นลำดับ แบ่งออกเป็น 4 ลำดับ ได้แก่ 1) งานวิจัยที่ตีพิมพ์ไม่เกิน 5 ปี 2) พิจารณาชื่อวรรณกรรมตรงมาหัวข้อ ข่าวปลอมปัญญาประดิษฐ์ และการประมวลผลภาษาธรรมชาติ 3) วรรณกรรมอยู่ในรูปแบบภาษาอังกฤษ 4) พิจารณาบทคัดย่อวรรณกรรม โดยเกณฑ์ในการพิจารณาแสดงถึงการตรวจสอบข่าวปลอมด้วยกระบวนการปัญญาประดิษฐ์ และพิจารณาวรรณกรรมฉบับเต็ม จึงทำการศึกษาจากงานวิจัยที่ทำด้านการตรวจสอบข่าวปลอมด้วยปัญญาประดิษฐ์ จำนวน 10 เรื่อง

ผลการวิจัย: ปัญญาประดิษฐ์สามารถตรวจสอบข่าวปลอมได้ถูกต้องเป็นที่น่าพอใจสูงที่มากกว่าร้อยละ 90 และได้แนวทางในการใช้ปัญญาประดิษฐ์ตรวจสอบข่าวปลอม แต่อย่างไรก็ตาม มีข้อสังเกตในการวิเคราะห์ข้อมูลที่ใช้ในการใช้สอนเครื่องคอมพิวเตอร์นั้น จะเหมาะสมกับแบบจำลองนั้น ๆ เสมอ แต่บนสื่อสังคมออนไลน์มีข่าวสารเกิดขึ้นทุกวินาที จึงจำเป็นต้องศึกษาพัฒนาปัญญาประดิษฐ์ตรวจสอบข่าวปลอม เพิ่มเติมเพื่อการตรวจสอบข่าวปลอมด้วยปัญญาประดิษฐ์ที่ทันเหตุการณ์

สรุปผล: ปัญญาประดิษฐ์สามารถตรวจสอบข่าวปลอมอย่างแม่นยำเกิน 90% และมีแนวทางในการใช้ปัญญาประดิษฐ์ในการตรวจสอบข่าวปลอม อย่างไรก็ตาม การใช้ข้อมูลในการสอนเครื่องคอมพิวเตอร์ต้องเป็นไปตามลักษณะของแบบจำลอง และเนื่องจากข่าวสารบนสื่อสังคมออนไลน์เป็นเหตุการณ์ทุกวินาที การพัฒนาปัญญาประดิษฐ์ต่อเพื่อตรวจสอบข่าวปลอมที่ทันเหตุการณ์เป็นสิ่งจำเป็น

คำสำคัญ : ปัญญาประดิษฐ์; การประมวลผลภาษาธรรมชาติ; ข่าวปลอม

บทนำ

ปัญหาสำคัญในสังคมไทย การรับรู้ข่าวปลอมส่งผลกระทบต่อทำให้เกิดความเสียหายมีการแพร่กระจายไปในวงกว้าง จึงมีความพยายามในการตรวจสอบควบคุมข่าวปลอม ทั้งด้านกฎหมายควบคุมการเผยแพร่ข้อมูลอันเป็นเท็จ (ส.อ.ท.สพธอ, 2016) ตัวอย่างการตรวจสอบข้อเท็จจริงที่มีการเผยแพร่ อาทิเช่น รายการข่าวก่อนแชร์ ซึ่งเป็น การตรวจสอบข้อเท็จจริงของข่าวปลอมโดยส่งผู้สื่อข่าวออกไปสัมภาษณ์ผู้ที่มีความรู้ความเชี่ยวชาญ เกี่ยวกับประเด็นที่กล่าวถึงในข่าวปลอม แล้วจึงนำมาเผยแพร่เปิดเผยข้อเท็จจริง นอกจากนี้ยังมีเว็บไซต์ที่พยายามตรวจสอบข่าวปลอมที่เกิดขึ้นทั้งต่างประเทศ และในประเทศไทย (Panyachot, 2021) เนื้อหาข้อมูลที่เป็นเท็จ หลอกลวง หรือข่าวที่สร้างสถานการณ์ รวมถึงการเขียนหรือการนำเสนอข่าวที่ได้รับการสนับสนุนอย่างปิดบังหรือแอบแฝง โดยมีความประสงค์เพื่อชักนำบุคคลอื่นให้มีความเชื่อในทางที่ผิดบิดเบือน และส่งผลให้เกิดความเสียหายต่อบุคคล หรือหน่วยงาน หากดูสถิติข่าวปลอมของประเทศไทย กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม (ดีอีเอส) เผยสถานการณ์ข่าวปลอมปี 2565 พบข่าวปลอมมุ่งบิดเบือนเกี่ยวกับนโยบายรัฐบาลและข่าวสารราชการมาแรงทั้งปี รองลงมาด้านสุขภาพและเตือนภัยพิบัติ กระทรวงดีอีเอสเดินทางขับเคลื่อนนโยบายการแก้ไขปัญหาข่าวปลอมอย่างต่อเนื่อง โดยเป็นการทำงานบูรณาการร่วมกับทุกหน่วยงานราชการ ในการรับแจ้งเบาะแสจากประชาชนผ่านช่องทางต่าง ๆ ของศูนย์ต่อต้านข่าวปลอมในช่องทางเพจ Facebook Twitter และ Line Official และติดตามการสนทนาบนโลกออนไลน์ (Social Listening) เกี่ยวกับข่าวปลอม โดยในรอบปี 2565 ภาพรวมผลการดำเนินการของศูนย์ต่อต้านข่าวปลอมตั้งแต่วันที่ 1 ม.ค.2565 – 31 ธ.ค. 2565 จากการรับแจ้งเบาะแสและติดตามการสนทนาบนโลกออนไลน์ พบว่ามีข้อความข่าวที่ต้องคัดกรอง 517,965,417 ข้อความ หลังจากคัดกรองพบข้อความข่าวที่ต้องดำเนินการตรวจสอบ 14,859 เรื่อง แบ่งเป็น โดยหมวดหมู่ นโยบายรัฐ 3,772 เรื่อง ตามมาด้วย หมวดหมู่สุขภาพ 2,344 เรื่อง หมวดหมู่เศรษฐกิจ 634 เรื่อง และหมวดหมู่ภัยพิบัติ 450 เรื่อง

จากปัญหาดังกล่าว การวิจัยครั้งนี้มุ่งเน้นไปที่แนวทาง วิธีดำเนินการวิจัย แบบจำลอง โมเดล ผลการวิจัย ในการจำแนกข่าวปลอมซึ่งเป็นพื้นฐานในการเป็นผู้รับสารในการพิจารณาสารเบื้องต้นเพื่อยกระดับการรู้เท่าทันสื่อให้มีประสิทธิภาพมากยิ่งขึ้น และเพื่อเป็นแนวทางในการใช้ปัญญาประดิษฐ์ตรวจสอบข่าวปลอม โดยผลลัพธ์จากการศึกษาวรรณกรรมฉบับเต็มจะถูกนำมาจัดหมวดหมู่และเปรียบเทียบผลการศึกษา ในแต่ละด้าน เพื่อให้เกิดเป็นข้อสรุปขององค์ความรู้ที่มีคุณภาพและเป็นแนวทางในการดำเนินงานวิจัยต่อไป ในอนาคต

วัตถุประสงค์การวิจัย

1. เพื่อศึกษาวิธีการและแบบจำลองการตรวจสอบข่าวปลอมด้วยปัญญาประดิษฐ์
2. เพื่อเสนอแนวทางการใช้ปัญญาประดิษฐ์ตรวจสอบข่าวปลอม

การทบทวนวรรณกรรม

ในการทบทวนวรรณกรรมเบื้องต้น ประกอบด้วยหัวข้อที่สำคัญ ดังนี้ 1.ข่าวปลอม (Fake News) 2. ปัญญาประดิษฐ์ (Artificial Intelligence: AI)

1. แนวคิดเกี่ยวกับข่าวปลอม

พจนานุกรมศัพท์นิเทศศาสตร์ฉบับราชบัณฑิตยสภาให้หมายความว่า ข่าวลวง: ข้อมูลข่าวสารที่ไม่ตรงกับความจริง แพร่กระจายแฝงในรูปแบบการรายงาน ข่าวมักใช้ภาษาเร้าอารมณ์ โดยมีวัตถุประสงค์ให้เชื่อ ข่าวปลอม: ข่าวที่มีเนื้อหาเป็นเท็จ ปรากฏจากข้อเท็จจริง ไม่สามารถตรวจสอบแหล่งที่มาได้ หรืออาจมีเจตนาบิดเบือนข้อมูลไปจากข้อเท็จจริงเดิม ข่าวปลอม ไม่ได้เป็นเรื่องใหม่ในสังคมไทย และสังคมโลก (bbc, 2017) การบิดเบือนข้อมูลข่าวสาร และการสร้างข่าวสารเพื่อโน้มน้าว หรือจูงใจคนให้เชื่อในวาทกรรมของผู้มีอำนาจนั้นเกิดขึ้นมานานแล้ว แต่เทคนิคที่เราเข้าใจในปัจจุบันส่วนใหญ่จะเป็นข้อมูลข่าวสารบนโลกออนไลน์ที่เป็นกระแส ถูกแชร์กันเป็นจำนวนมาก มีเนื้อหาที่ไม่ถูกต้องตามความเป็นจริง แต่ถูกแชร์เพื่อสร้างความตื่นตระหนก (Nichols, 2020) ข้อมูลบางประเภทสร้างขึ้นเพื่อให้ผู้รับสื่อมีความเชื่อด้านใดด้านหนึ่งอย่างสุดโต่ง จนไม่ยอมรับข้อมูลอีกด้าน ส่งผลให้เกิดความขัดแย้งกันในสังคม เนื่องจากผู้รับข้อมูลข่าวสารเหล่านี้จะอยู่ในกลุ่มของผู้ที่คิดเห็นเหมือนๆ กัน และจะเชื่อว่าสิ่งที่ตนคิดถูกต้อง อีกฝ่ายเป็นข้อมูลที่ผิดที่ไม่ถูกต้อง ซึ่งอันตรายต่อการมีส่วนร่วมในกระบวนการทางประชาธิปไตย งานศึกษาเรื่อง ข่าวลวง: ปัญหาและความท้าทาย ของ (เฉลิมชัย กิกเกียรติกุล & ธีรนนทณัฐ ตานไพบุลย์, 2018) ได้กล่าวไว้ว่าตอนหนึ่งว่า คณะกรรมาธิการยุโรป (European Commission) ให้นิยามข่าวลวง หมายถึงข้อมูลชุดหนึ่งที่ทำขึ้นมาโดยมีเจตนาจะบิดเบือนและโจมตีบุคคล กลุ่มคน หรือองค์กร และบางครั้งข่าวลวงหมายความรวมถึงการเขียนข่าวเชิงลบ การโฆษณาชวนเชื่อทางการเมือง และรูปแบบโฆษณาที่ชวนให้เข้าใจผิดจากทัศนคติของผู้รับสาร รวมถึงการรายงานข่าวในรูปแบบเสียดสีและมีการให้ข้อมูลที่ผิด โดยเจตนาโจมตีทางไซเบอร์ เมื่อนิยามอย่างแคบ ข่าวลวงเป็นข่าวที่ถูกสร้างขึ้นในการรายงานข่าว ซึ่งรวมถึงสำนักข่าว นักการเมือง ข่าวลวงในบริบทของการเมืองสหรัฐอเมริกา (Andrea Carson, 2019) มีการหยิบยกมาอภิปรายอย่างแพร่หลายในช่วงการเลือกตั้งประธานาธิบดี ค.ศ. 2016 พรรคเดโมแครต และรีพับลิกันต่างกล่าวหาซึ่งกันและกันและกล่าวหาว่าสื่อมวลชน ที่ไม่เป็นกลาง เข้าข้างอีกพรรคหนึ่งก่อกวนเท็จขึ้นมาใส่ร้าย ดังนั้นข่าวลวงในบริบทของการเมืองและการหาเสียงเลือกตั้งกลายเป็นข่าวโกหกและใส่ร้ายซึ่งเป็นส่วนหนึ่งของวัฒนธรรมทางการเมืองและดูเหมือนจะแพร่กระจายไปอย่างรวดเร็วและควบคุมได้ยาก เนื่องจากมีการใช้สื่อใหม่เป็นเครื่องมือในการแพร่กระจาย ความผิดปกติของข้อมูลข่าวสาร (Information Disorder) (Derakhshan, 2018) จะสามารถแบ่งตามฐานข้อมูลได้ 2 ส่วน คือ ข้อมูลที่ไม่ถูกต้อง (Falseness) และข้อมูลที่มีเจตนามุ่งร้าย (Intent to harm) และถูกแบ่งออกลงไปอีก 3 กลุ่ม คือ ข้อมูลที่ผิด (Misinformation), ข้อมูลที่บิดเบือน (Disinformation) และข้อมูลที่แฝงเจตนาร้าย (Misinformation) ซึ่งในแต่ละกลุ่มก็จะมีประเภทของข้อมูลแยกย่อยลงไปอีกดังนี้

1.1 ข้อมูลที่ผิด (Misinformation) ข้อมูลที่ผิดพลาดหรือข้อมูลเท็จ เป็นข้อความหรือรูปภาพที่ไม่ได้เกิดจากเจตนาต้องการเผยแพร่ข้อมูลที่ผิด หรือตั้งใจบิดเบือนข้อมูล แต่เป็นกรณีที่ผู้รับสารเผยแพร่ไปด้วยความเข้าใจและมีความเชื่อว่าข้อมูลเหล่านั้นถูกต้อง จึงเกิดความผิดพลาด อาทิเช่น การเสียดสีหรือล้อเลียน (Satire or Parody) ถึงแม้ว่าการเสียดสีหรือการล้อเลียนจะถือเป็นศิลปะรูปแบบหนึ่ง แต่ก็ยังเป็นสิ่งที่นำไปสู่การเผยแพร่ข้อมูลที่ผิดหรือข้อมูลที่บิดเบือนได้ ข้อมูลที่มีการเชื่อมโยงผิด (False Connection) หัวเรื่องรูปภาพ วิดีโอ หรือแคปชั่นที่ประกอบเนื้อหาไม่เป็นไปในทางเดียวกันกับเนื้อหา เช่น เนื้อหาประเภท Clickbait ที่เมื่อกดเข้าไปอ่านแล้ว ทำให้ผู้รับสารรู้สึกเหมือนถูกหลอก และเนื้อหาซึ่งนำไปสู่การเข้าใจผิด

(Misleading Content) เนื้อหาใช้การตัดบางส่วนของข้อมูลมาเสนอ ทั้งตั้งใจและไม่ตั้งใจจนอาจก่อให้เกิดความเข้าใจผิดในข้อมูล เช่น ตัดบางคำพูด ซึ่งอาจทำให้เข้าใจผิด เมื่อเปรียบเทียบกับเนื้อหาฉบับเต็ม

1.2 ข้อมูลที่บิดเบือน (Disinformation) คือข้อมูลที่ผิด แต่ผู้รับสารเผยแพร่ไปด้วยความตั้งใจที่รู้ว่ามีข้อมูลเหล่านั้นไม่ถูกต้อง หรือก็มีเจตนาที่จะบิดเบือนข้อมูลและชี้นำผู้คนที่เกิดความเข้าใจผิด เช่น เนื้อหาที่แต่งขึ้นมาใหม่ (Fabricated Content) ข้อมูลที่ใหม่หาแหล่งข้อมูลสนับสนุนเพิ่มเติมไม่ได้ อาจเป็นข้อมูลที่แต่งขึ้นมาใหม่นำไปสู่ข้อมูลผิด รวมถึงการบิดเบือนข้อมูลตรวจสอบได้ยาก จึงต้องใช้การวิเคราะห์ แยกแยะ และตรวจสอบให้รอบครอบ เนื้อหาที่มีการดัดแปลง (Manipulated Content) ภาพหรือข้อมูลจริงถูกนำมาดัดแปลงจากต้นฉบับ มีเจตนาเพื่อบิดเบือนข้อมูล หลอกลวง ชี้นำให้เกิดความเข้าใจผิด เช่น การตัดต่อภาพ เสียง หรือวิดีโอ ข้อมูลที่ผิดบริบท (False Context) รูปแบบการนำเสนอในรูปแบบต่าง ๆ ที่ถูกนำมาหมุนเวียนซ้ำจากบริบทเดิมเปลี่ยนไปเรื่อย ๆ เช่น นำภาพจากเหตุการณ์หนึ่งมาใช้กับอีกเหตุการณ์หนึ่ง ทำให้เกิดความเข้าใจผิด เป็นต้น และข้อมูลที่มีเนื้อหาแอบอ้าง (Imposter Content) การแอบอ้างหรือสวมรอยเป็นสื่อที่น่าเชื่อถือเผยแพร่ข้อมูลข่าวสาร หากผู้รับสารไม่ได้ตรวจสอบอาจเกิดความเข้าใจผิด จนเกิดการส่งต่อข่าวสาร เนื่องจากคิดว่าเป็นข้อมูลจากสื่อหรือองค์กรที่น่าเชื่อถือและส่งผลกระทบต่อภาพลักษณ์สื่อหรือองค์กรได้

1.3 ข้อมูลที่แฝงเจตนาร้าย (Misinformation) อาจเป็นข้อมูลที่มาจากข้อเท็จจริง แต่ถูกใช้เป็นเครื่องมือในการทำร้ายผู้อื่น เช่น ข้อมูลที่ละเมิดความเป็นส่วนตัว ข้อมูลที่ขัดต่อมาตรฐานและจรรยาบรรณสื่อ เช่น ภาพเปลือย รวมถึงการคุกคาม (Harassment) หรือข้อความหรือคำพูดที่แสดงความเกลียดชัง (Hate speech) เป็นต้น สังเกตได้ว่าความผิดปกติของข้อมูลข่าวสารแต่ละประเภทก็มีความคาบเกี่ยวกันว่าจะเป็นข้อมูลที่ผิด (Misinformation), ข้อมูลที่บิดเบือน (Disinformation) หรือข้อมูลที่แฝงเจตนาร้าย (Misinformation) ทั้งนี้ขึ้นอยู่กับเจตนา ข้อมูล และบริบทต่าง ๆ

2.แนวคิดเกี่ยวกับปัญญาประดิษฐ์ (Artificial Intelligence: AI)

ปัญญาประดิษฐ์ (AI) คือเครื่องจักร (machine) ที่มีฟังก์ชันที่มีความสามารถในการทำความเข้าใจ เรียนรู้องค์ความรู้ต่าง ๆ อาทิเช่น การรับรู้ การเรียนรู้ การให้เหตุผล และการแก้ปัญหา ต่าง ๆ เครื่องจักรที่มีความสามารถเหล่านี้ก็ถือว่าเป็น ปัญญาประดิษฐ์ เพราะฉะนั้นจึงสามารถกล่าวได้ว่า AI ถือกำเนิดขึ้นเมื่อเครื่องจักรมีความสามารถที่จะเรียนรู้ตัวเอง ซึ่ง AI ก็ถูกแบ่งออกเป็นหลายระดับตามความสามารถหรือความฉลาด โดยจะวัดจากความสามารถในการให้เหตุผล การพูด และทัศนคติ สรุปสิ่งที่มนุษย์สร้างขึ้น ให้มีความสามารถในการคิดได้เองโดยไม่ต้องสอน คิดและแสดงออกแบบมีเหตุผลได้เหมือนมนุษย์ คือการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) การวิเคราะห์ข้อมูลต่าง ๆ และการสร้างแบบจำลองจากข้อมูลเหล่านั้นด้วยศาสตร์ทางการเรียนรู้ของเครื่อง (Machine Learning) จะเกี่ยวข้องกับข้อมูลที่มีโครงสร้าง (Structured Data) เช่น ข้อมูลเชิงตาราง (Tabular Data) เป็นส่วนใหญ่ แต่ในโลกความเป็นจริง (Harbert, 2021) ข้อมูลบนโลกดิจิทัลในปัจจุบันมากกว่า 80% เป็นข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) เช่น ไฟล์ภาพ เสียง หรือ วิดีโอ รวมถึงข้อมูลในลักษณะของข้อความ อาทิ เนื้อหาในบทความต่าง ๆ การโพสต์ข้อความบนโซเชียลมีเดีย บทสัมภาษณ์ของนักการเมือง และการแสดงความคิดเห็นต่อสินค้าบนแพลตฟอร์มพาณิชย์อิเล็กทรอนิกส์ (E-commerce Platform) ทั้งนี้การวิเคราะห์ข้อมูลในลักษณะดังกล่าวจำเป็นต้องอาศัยศาสตร์เฉพาะทางคือ การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) เพื่อช่วยให้คอมพิวเตอร์เข้าใจภาษาที่ซับซ้อนของมนุษย์ วิวัฒนาการและความสำคัญของ

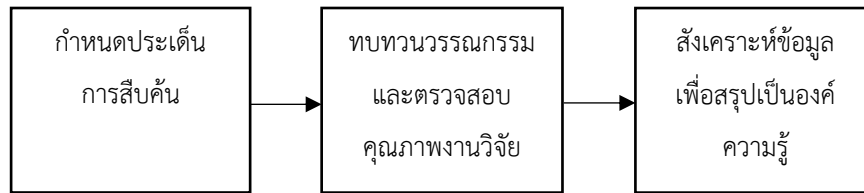
การประมวลผลภาษาธรรมชาติ (NLP) เป็นศาสตร์ที่สำคัญทางด้าน Machine Learning โดยเป็นสาขาวิชาหนึ่งที่ประกอบด้วยองค์ความรู้จากหลากหลายแขนง อาทิ ภาษาศาสตร์ (Linguistics) วิทยาการคอมพิวเตอร์ (Computer Science) ปัญญาประดิษฐ์ (Artificial Intelligence: AI) รวมถึงสถิติ (Statistics) โดยมีจุดมุ่งหมายเพื่อให้คอมพิวเตอร์สามารถทำความเข้าใจ ข้อมูลที่มีลักษณะเป็นข้อความหรือคำพูด เช่นเดียวกับที่มนุษย์ที่ทำ จุดกำเนิดมาตั้งแต่ช่วงกลางศตวรรษที่ 19 และได้มีการพัฒนาต่อยอดมาเรื่อย ๆ จนถึงปัจจุบัน ด้วยพลังการคำนวณของคอมพิวเตอร์ที่มีการพัฒนาสูงขึ้นอย่างต่อเนื่อง ทำให้เทคโนโลยีที่มีความซับซ้อนสูงอย่าง การเรียนรู้เชิงลึก (Deep Learning) ถูกนำมาใช้งานแทนที่ Machine Learning ซึ่งใช้ความรู้ทางด้านสถิติแบบดั้งเดิม อย่างแพร่หลายมากขึ้น รวมถึงในงานด้าน NLP ด้วยเช่นกัน อาทิ การสร้างแบบจำลองทางภาษา (Language Model) และการวิเคราะห์โครงสร้างของข้อความ (Parsing) NLP ได้รับความสนใจอย่างแพร่หลายและมีการพัฒนาอย่างต่อเนื่องมาตลอดหลายสิบปี คือความต้องการในการประมวลผลข้อมูลที่มีลักษณะเป็นข้อความในหลายภาคส่วน อาทิ ด้านเทคโนโลยีการสื่อสาร ด้านการศึกษา และด้านธุรกิจ ล้วนแล้วแต่มีการป้อนข้อมูลดังกล่าวเข้าสู่โลกดิจิทัลเป็นปริมาณมหาศาล ดังนั้นหากไม่มีกระบวนการที่เหมาะสมในการนำข้อมูลมาใช้ประโยชน์ เครื่องมือเบื้องต้นที่น่าสนใจสำหรับงานการประมวลผลภาษาธรรมชาติ

2.1 Tokenization คือ กระบวนการตัดคำออกจากข้อความยาว ๆ ซึ่งแต่ละคำที่ถูกตัดออกมาจะถูกเรียกว่า Token โดยกระบวนการนี้มักจะถูกทำเป็นกระบวนการแรก และ Token

2.2 Word Embedding คือ กระบวนการในการแปลงคำให้เป็นรูปแบบเชิงปริมาณอย่างเช่น ตัวเลข หรือ เวกเตอร์ ซึ่งการทำให้เป็นเชิงปริมาณจะช่วยให้ง่ายต่อการนำไปใช้ในการคำนวณต่าง ๆ การทำ Word Embedding เช่น 1) One-hot Encoding: (Chen, 2020) การแปลงคำศัพท์ให้กลายเป็นเวกเตอร์ที่มีเฉพาะเลข 0 และ 1 ซึ่งมีขนาดเท่ากับจำนวนของคำทั้งหมด 2) Term Frequency – Inverse Document Frequency (TF-IDF): (Prasertsom, 2020) การให้ค่าความสำคัญของคำผ่านการพิจารณาองค์ประกอบของคำในบทความ โดยคำที่ปรากฏบ่อยในบทความหนึ่ง ๆ แต่ปรากฏได้น้อยในบทความอื่นจะมีค่า TF-IDF ที่สูง 3) Word2Vec: (Prasertsom, P., 2021) การสร้างเวกเตอร์ของคำโดยพิจารณาบริบทของคำนั้นตามคำที่อยู่ข้างเคียง 4) Parsing คือ กระบวนการในการวิเคราะห์โครงสร้างของข้อความ โดยเป็นการตรวจสอบว่าคำที่เป็นส่วนประกอบนั้นมีความถูกต้องตามหลักไวยากรณ์ของภาษาหรือไม่ ซึ่งผลการวิเคราะห์จะออกมาเป็นโครงสร้างแบบต้นไม้ ที่เรียกว่า Parse Tree 5) Part of Speech Tagging (PoS Tagging) คือ กระบวนการระบุประเภทของคำตามหน้าที่ในประโยค อาทิ คำนาม (Nouns) คำสรรพนาม (Pronouns) คำคุณศัพท์ (Adjectives) คำกริยา (Verbs) คำกริยาวิเศษณ์ (Adverbs) คำบุพบท (Prepositions) คำสันธาน (Conjunctions) และ คำอุทาน (Interjections) 6) Named Entity Recognition (NER) คือ กระบวนการในการระบุหมวดหมู่รวมถึงตำแหน่งของคำในเอกสาร เช่น ชื่อคน ชื่อองค์กร สถานที่ วันเวลา จำนวนเงิน 7) Language Model หรือ แบบจำลองทางภาษา คือ แบบจำลองทางสถิติจากคลังข้อมูลของคำในภาษาเพื่อแสดงถึงความน่าจะเป็นของลำดับของคำ โดยหนึ่งในการประยุกต์ใช้แบบจำลองทางภาษาที่คุ้นเคย ได้แก่ ระบบการคาดเดาและแนะนำคำ เช่น การแนะนำคำถัดไปในช่องค้นหาของ Google รวมถึงการค้นหาในแพลตฟอร์ม E-commerce สำหรับในปัจจุบัน แบบจำลองทางภาษาได้ถูกนำเสนอออกอย่างหลากหลาย อาทิ BERT และ GPT-3 (Brown, 2020) ซึ่งได้รับการพัฒนาโดยบริษัท Google และ Open AI ตามลำดับ 8) Topic Model คือ แบบจำลองทางสถิติซึ่งสร้างมาจากกลุ่มของเอกสาร เพื่อวิเคราะห์ค้นหา

หัวข้อต่าง ๆ ที่มีการพูดถึงร่วมกันอยู่ในกลุ่มของเอกสารเหล่านั้น โดยผลลัพธ์ของแบบจำลองจะทำให้ทราบว่า จากปริมาณเอกสารจำนวนมากมีการพูดถึงเรื่องใดบ้าง ซึ่งจะช่วยให้ผู้ที่ไม่ต้องเสียเวลาอ่านเอกสารทั้งหมด 9) Sentiment Analysis คือ การวิเคราะห์ความรู้สึกที่แสดงออกผ่านทางข้อความ ซึ่งรวมถึงการวิเคราะห์อารมณ์ของผู้เขียนข้อความ

กรอบแนวคิดการวิจัย



แผนภาพที่ 1 กรอบแนวคิดการวิจัย

ระเบียบวิธีการวิจัย

การตรวจสอบข่าวปลอมด้วยปัญญาประดิษฐ์ เพื่อให้การทบทวนวรรณกรรมเป็นที่โปร่งใสและตรงตามหลักการทางวิทยาศาสตร์ซึ่งสามารถพิสูจน์ซ้ำได้ผู้วิจัยจึงนำกระบวนการในการทบทวนวรรณกรรมอย่างเป็นระบบตามรูปแบบของ Briner และ Denyer (Briner & Denyer, 2012) และการใช้คุณลักษณะบางส่วนจาก PRISMA Statement โดย David Moher (Moher et al., 2015) และคณะ มาประยุกต์ใช้กับงานวิจัยนี้ โดยแบ่งออกเป็น 3 ขั้นตอน

- 1) กำหนดประเด็นการสืบค้น
- 2) ทบทวนวรรณกรรมและตรวจสอบคุณภาพงานวิจัย
- 3) สังเคราะห์ข้อมูลเพื่อสรุปเป็นองค์ความรู้

1.กำหนดประเด็นการสืบค้น การสืบค้นวรรณกรรมเพื่อนำมาใช้ในการทบทวนวรรณกรรมอย่างเป็นระบบเริ่มต้นขึ้นในช่วงปี มกราคม 2561 ถึง พฤษภาคม 2566 ผู้วิจัยค้นหาวรรณกรรมทางวิชาการ ซึ่งได้มีการกำหนดคำค้นเป็นภาษาอังกฤษแบ่งออกเป็น 2 คำค้น ได้แก่ “Fake News”, “Artificial Intelligence: AI” และ “Natural Language Processing: NLP” ผู้วิจัยคัดเลือกเบื้องต้นเฉพาะวรรณกรรมที่ผ่านการถูกอ้างอิงมาแล้วเท่านั้น โดยแหล่งที่มาของวรรณกรรมจะต้องเป็นที่ยอมรับและมีความน่าเชื่อถือ ต้องเป็นรูปแบบภาษาอังกฤษ ฐานข้อมูล วารสารเข้าถึงแบบเปิด (Open Access Journal) เพราะนักวิจัยและบุคคลที่สนใจสามารถค้นหาได้แบบไม่มีข้อจำกัด การคัดกรองวรรณกรรมจะถูกพิจารณาเป็นลำดับ โดยเริ่มจากการพิจารณาชื่อวรรณกรรม บทคัดย่อวรรณกรรม และวรรณกรรมฉบับเต็มตามลำดับ หากพบว่าลำดับใดไม่ตรงตามเกณฑ์ที่ตั้งไว้วรรณกรรมนั้นจะถูกลำเอียง

2.ทบทวนวรรณกรรมและตรวจสอบคุณภาพงานวิจัย วรรณกรรมที่ถูกลำเอียงเข้าสู่การพิจารณาเป็นลำดับ จำเป็นต้องผ่านการประเมินจากผู้วิจัยตามเกณฑ์ที่ผู้วิจัย กำหนดเพื่อคัดเลือกวรรณกรรมที่มีคุณภาพ โดยเกณฑ์ที่ ผู้วิจัยกำหนดจะแบ่งการพิจารณาออกเป็น 2 ส่วน ได้แก่ เกณฑ์การคัดเลือกผลงาน (Inclusion Criteria) และเกณฑ์การคัดผลงานออก (Exclusion Criteria)

ตารางที่ 1 เกณฑ์การคัดเลือกผลงานเข้าเกณฑ์

เกณฑ์การคัดเลือก	Description
ผลงาน	
วันเวลา:	ที่ตีพิมพ์ไม่เกิน 5 ปี ตั้งแต่ มกราคม 2561 ถึง พฤษภาคม 2566
คำสำคัญ:	Fake News, Artificial Intelligence: AI และ Natural Language Processing: NLP
ภาษา:	อังกฤษ
วิธีดำเนินการวิจัย:	การตรวจสอบข่าวปลอมด้วยกระบวนการปัญญาประดิษฐ์

ซึ่งเกณฑ์ในการพิจารณาวรรณกรรมจะถูกพิจารณาเป็นลำดับ แบ่งออกเป็น 4 ลำดับ ได้แก่

- 1) งานวิจัยที่ตีพิมพ์ไม่เกิน 5 ปี ตั้งแต่ มกราคม 2561 ถึง พฤษภาคม 2566
- 2) พิจารณาชื่อวรรณกรรม หัวข้อ ข่าวปลอม ปัญญาประดิษฐ์ และการประมวลภาษาธรรมชาติ
- 3) วรรณกรรมอยู่ในรูปแบบภาษาอังกฤษ
- 4) พิจารณาบทความวรรณกรรม โดยเกณฑ์ในการพิจารณาแสดงถึงการตรวจสอบข่าวปลอมด้วยกระบวนการปัญญาประดิษฐ์ และพิจารณาวรรณกรรมฉบับเต็ม

3.สังเคราะห์ข้อมูลเพื่อสรุปเป็นองค์ความรู้ จากขั้นตอนทบทวนวรรณกรรมและตรวจสอบคุณภาพงานวิจัยดังที่กล่าวมาข้างต้นทำให้ได้รายการวรรณกรรมที่ผ่านการคัดเลือก ผลงานและถือเป็นวรรณกรรมที่มีคุณภาพ ในขั้นตอนนี้ ผู้วิจัยได้ดำเนินการทบทวนวรรณกรรมฉบับเต็มโดยละเอียด เพื่อหาผลลัพธ์ที่มุ่งเน้นใน 2 ด้าน ได้แก่ (1) การตรวจสอบข่าวปลอมด้วยปัญญาประดิษฐ์ และ (2) แนวทางในการใช้ปัญญาประดิษฐ์ตรวจสอบข่าวปลอม โดยผลลัพธ์จากการศึกษาวรรณกรรมฉบับเต็มจะถูกนำมาจัดหมวดหมู่และเปรียบเทียบผลการศึกษา ในแต่ละด้าน เพื่อให้เกิดเป็นข้อสรุปขององค์ความรู้ที่มี คุณภาพและเป็นแนวทางในการดำเนินงานวิจัยต่อไป ในอนาคต

ตารางที่ 2 ขั้นตอนการเตรียมข้อมูลโดยใช้ปัญญาประดิษฐ์กับการตรวจสอบข่าวปลอม

การเตรียมข้อมูล	(Devarajan, Nagarajan, Amanullah, Mary, & Cueva et al., 2020)	(Mookdarsanit, P., 2021)	Ahmed, Hinkelmann, & Corradini, 2022	(Deepak Sa, 2020)	(Sedik, Abohany, Sallam, Munasinghe, & Medhat, 2022)	(Garg & Kumar Sharma, 2022)	(Jain, Shakya, Khatter, & Gupta, 2019)	(Kaliyar, Goswami, Narang, & Sinha, 2020)	(Jadhav & Thepade, 2019)
การเลือกข้อมูล (Selection)	✓	✓	✓	✓		✓	✓		
การเตรียมข้อมูล (Preprocessing)	✓	✓	✓	✓	✓	✓	✓	✓	
การเปลี่ยนรูปข้อมูล (Transformation)	✓	✓	✓	✓	✓	✓			

การเตรียมข้อมูล แบ่งออกเป็น 3 ส่วน 1.การเลือกข้อมูล(Selection) 2.การเตรียมข้อมูล (Preprocessing) 3.การเปลี่ยนรูปข้อมูล (Transformation) โดยงานวิจัยส่วนใหญ่มีการใช้กระบวนการเตรียมข้อมูล

การสอนเครื่องคอมพิวเตอร์ให้เข้าใจข้อมูล (train) การทดสอบการทำนาย (test) การเพิ่มประสิทธิภาพแบบเบย์ของ Scikit-Optimize ถูกนำมาใช้เพื่อค้นหาไฮเปอร์พารามิเตอร์ที่เหมาะสมที่สุด N-Grams ใช้เพื่อทำความเข้าใจว่า CNN จัดการกับภาษาธรรมชาติ การจัดทำข้อความภาษาไทย การแบ่งคำในภาษาไทย เทคนิคที่ใช้ในการขยายขนาดของชุดข้อมูลเป้าหมายโดยการรวมคำหรือวลีที่มีความหมายคล้ายกัน ใช้อัลกอริทึมการเรียนรู้ของเครื่องคือ Count Vectorizer, TF-IDF Vectorizer พร้อมด้วยการประมวลผลภาษาธรรมชาติ (NLP) ใช้โดเมนต้นทาง (source domain) และชื่อผู้แต่ง (author names) โมเดลนี้อ้างอิงจาก LSTM และ FF Neural Networks ประสิทธิภาพของการแสดงเวกเตอร์คำต่าง ๆ เช่น GloVe และ word2vec

ตารางที่ 3 ขั้นตอนการเลือกใช้แบบจำลอง / โมเดล โดยใช้ปัญญาประดิษฐ์กับการตรวจสอบข่าวปลอม

แบบจำลอง / โมเดล	(Devarajan et al., 2023)	(Cueva et al., 2020)	(Mookdarsanit, P., 2021)	Ahmed, Hinkelmann, & Corradini, 2022	(Deepak Sa, 2020)	(Sedik et al., 2022)	(Garg & Kumar Sharma, 2022)	(Jain et al., 2019)	(Kaliyar et al., 2020)	(Jadhav & Thepade, 2019)
Natural Language Processing (NLP)	✓	✓	✓	✓	✓					
Convolutional Neural Network (CNN)				✓		✓	✓	✓	✓	
Deep learning								✓		
Neural Networks		✓			✓	✓		✓		✓
Generative			✓					✓		

แบบจำลองที่ใช้ แบ่งข้อมูล ข้อมูลที่แยกออกมาจะได้รับการวิเคราะห์โดยใช้เทคนิค NLP เพื่อระบุรูปแบบและคุณสมบัติที่สามารถช่วยแยกแยะระหว่างข่าวจริงและข่าวปลอม แบ่งออกเป็นกลุ่ม ๆ ได้ 5 กลุ่ม 1.Natural Language Processing (NLP) 2.Convolutional Neural Network (CNN) 3.Deep learning 4.Neural Networks 5.Generative

การวัดประสิทธิภาพ	(Devarajan et al., 2023)	(Cueva et al., 2020)	(Mookdaisanit, P., 2021)	Ahmed, Hinkelmann, & Corradini, 2022	(Deepak Sa, 2020)	(Sedik et al., 2022)	(Garg & Kumar Sharma, 2022)	(Jain et al., 2019)	(Kaliyar et al., 2020)	(Jadhav & Thepade, 2019)
ร้อยละ 95 ขึ้นไป	✓					✓			✓	✓
ร้อยละ 90 ขึ้นไป				✓			✓	✓		
ร้อยละ 85 ขึ้นไป										
ร้อยละ 80 ขึ้นไป										

การวัดประสิทธิภาพ เรื่องที่ 1 เฉลี่ย 99.72% และคะแนน F1 ที่ 98.33% ซึ่งมีประสิทธิภาพดีกว่าวิธีการอื่นๆ ที่มีอยู่ เรื่องที่ 2 โมเดล NLP เป็นโมเดลที่สามารถระบุการล้อเลียนว่าเป็นชาวปลอมได้ และยังสามารถปรับปรุงแบบจำลอง LSTM และ GRU เพื่อให้ NLP สามารถระบุได้ดียิ่งขึ้น เรื่องที่ 3 วิธีการตรวจจับชาวปลอมข้อความภาษาไทยที่เสนอโดยใช้โมเดล Thai-NLP และ Feature shifting มีประสิทธิภาพดีที่สุดคือโมเดล ULMFIT เรื่องที่ 4 ใช้อัลกอริทึมการเรียนรู้ของเครื่องและการประมวลผลภาษาธรรมชาติ (NLP) การจำแนกชาวปลอมและระบุคุณสมบัติหลักที่สามารถใช้ในการตรวจจับชาวปลอม ระบบที่พัฒนาขึ้นมีความแม่นยำสูงถึง 93% เรื่องที่ 5 โดเมนต้นทางและชื่อผู้เขียนเพื่อเลียนแบบกระบวนการตรวจสอบข้อเท็จจริง เรื่องที่ 6 ได้รับความถูกต้องของ 99.6% และการ train ใช้อัลกอริทึมได้เร็วที่สุด เรื่องที่ 7 มีความแม่นยำถึงสูงสุดที่ 93.8% เรื่องที่ 8 แบบจำลองที่นำเสนอมีความแม่นยำ 93.6% เรื่องที่ 9 ความถูกต้อง 98.36% 10.DSSM และ RNN ที่ปรับปรุงแล้ว ความแม่นยำถึง 99% ปัญญาประดิษฐ์ (AI) สามารถตรวจสอบชาวปลอมได้ถูกต้องเป็นที่น่าพอใจสูงที่มากกว่าร้อยละ 90 และได้แนวทางในการใช้ปัญญาประดิษฐ์ตรวจสอบชาวปลอม

อภิปรายผล

การตรวจสอบชาวปลอมด้วยปัญญาประดิษฐ์ เพื่อวิเคราะห์และเปรียบเทียบคุณสมบัติเฉพาะของเทคโนโลยีปัญญาประดิษฐ์ (AI), การเรียนรู้ของเครื่อง (ML), การประมวลผลภาษาธรรมชาติ (NLP) ด้านปัญญาประดิษฐ์ สามารถที่จะตรวจสอบชาวปลอมได้ ส่วนลักษณะของชาวปลอมก็เป็นคุณลักษณะเฉพาะที่สามารถประยุกต์ข้อมูลเพื่อให้เทคโนโลยีปัญญาประดิษฐ์สามารถใช้เทคนิคต่าง ๆ เช่น รวบรวมจากแพลตฟอร์มโซเชียลมีเดีย จัดทำความสะดวกข้อมูล แบ่งคำ วลี ประโยค แยกคุณลักษณะ การจำแนก (Classification) การวิเคราะห์การจัดกลุ่ม (Cluster Analysis) การพยากรณ์ (Prediction) และวัดประสิทธิภาพ การศึกษาจากงานวิจัยที่ทำด้านการตรวจสอบชาวปลอมด้วยปัญญาประดิษฐ์ จำนวน 10 เรื่อง



ตรงตามวัตถุประสงค์ที่ 1) เพื่อศึกษาวิธีการและแบบจำลองการตรวจสอบข่าวปลอมด้วยปัญญาประดิษฐ์ และตรงตามวัตถุประสงค์ที่ 2) เพื่อเสนอแนวทางการใช้ปัญญาประดิษฐ์ตรวจสอบข่าวปลอม จากงานวิจัยที่รีวิวนสามารถประยุกต์หรือเสนอแนวทางการตรวจสอบเป็นรายด้าน ดังนี้ การเตรียมข้อมูล แบ่งออกเป็น 3 ส่วน 1.การเลือกข้อมูล (Selection) 2.การเตรียมข้อมูล (Preprocessing) 3.การเปลี่ยนรูปข้อมูล (Transformation) โดยงานวิจัยส่วนใหญ่มีการใช้กระบวนการเตรียมข้อมูลการสอนเครื่องคอมพิวเตอร์ให้เข้าใจข้อมูล (train) การทดสอบการทำนาย (test) การเพิ่มประสิทธิภาพแบบเบย์ของ Scikit-Optimize เทคนิคที่ใช้ในการขยายขนาดของชุดข้อมูลเป้าหมายโดยการรวมคำหรือวลีที่มีความหมายคล้ายกัน ใช้อัลกอริทึมการเรียนรู้ของเครื่องคือ Count Vectorizer, TF-IDF Vectorizer สามารถใช้โดเมนต้นทาง (source domain) และชื่อผู้แต่ง (author names) ประสิทธิภาพของการแสดงเวกเตอร์ค่าต่าง ๆ เช่น GloVe และ word2vec แบบจำลองที่ใช้ แบ่งข้อมูล ข้อมูลจะได้รับการวิเคราะห์โดยใช้เทคนิค แบ่งออกเป็นกลุ่ม ๆ ได้ 5 กลุ่ม 1.Natural Language Processing (NLP) 2.Convolutional Neural Network (CNN) 3.Deep learning 4.Neural Networks 5.Generative การวัดประสิทธิภาพปัญญาประดิษฐ์ (AI) สามารถตรวจสอบข่าวปลอมได้ถูกต้องเป็นที่น่าพอใจสูงที่มากกว่าร้อยละ 90 และได้แนวทางในการใช้ปัญญาประดิษฐ์ตรวจสอบข่าวปลอม

ข้อเสนอแนะ

การตรวจสอบข่าวปลอมด้วยปัญญาประดิษฐ์ เพื่อวิเคราะห์และเปรียบเทียบคุณสมบัติเฉพาะของเทคโนโลยีปัญญาประดิษฐ์ การเรียนรู้ของเครื่อง การประมวลผลภาษาธรรมชาติ โครงข่ายประสาทเทียม แต่รูปแบบการเรียนรู้ของเครื่องอาจมีความเหมือนหรือคล้ายกับข้อมูลแต่ละลักษณะด้านปัญญาประดิษฐ์ และข้อมูลที่ใช้ในการใช้สอนเครื่องคอมพิวเตอร์นั้น จะเหมาะสมกับแบบจำลองนั้น ๆ เสมอ แต่ข้อมูลมีการเพิ่มขึ้นบนสื่อสังคมออนไลน์ทุกวินาทีเป็นจำนวนมาก จึงจำเป็นที่จะต้องศึกษาวิจัยวิธีการตรวจสอบข่าวปลอมด้วยปัญญาประดิษฐ์ เพื่อการตรวจสอบข่าวปลอมที่ทันเหตุการณ์ตามการสื่อสารผ่านเครือข่ายอินเทอร์เน็ต

การตรวจสอบข่าวปลอมแบบ Realtime ในความเป็นจริงทำได้ยาก แต่สามารถตรวจสอบแหล่งที่มาที่ถูกต้องก่อนได้ จากนั้นจึงต้องอาศัยข้อมูลที่ train หรือข้อมูลจากผู้เชี่ยวชาญมายืนยันผลลัพธ์อีกครั้ง

เอกสารอ้างอิง

เฉลิมชัย ก๊กเกียรติกุล และ ธัญชนกนัฐ ด่านไพบูลย์. (2018). Fake News: Problems and Challenge.

NBTC Journal, 2(2), 173 - 192. Retrieved from [https://so04.tci-](https://so04.tci-thaijo.org/index.php/NBTC_Journal/article/view/146058)

[thaijo.org/index.php/NBTC_Journal/article/view/146058](https://so04.tci-thaijo.org/index.php/NBTC_Journal/article/view/146058)

ส.อ.ห.สพธอ. (2016). *รู้จัก พ.ร.บ. คอมพิวเตอร์ฯ ฉบับที่ 2 พ.ศ. 2560*. Retrieved from:

<https://www.etda.or.th/th/newsevents/pr-news/ICT-LAW/รจก-พ-ร-บ-คอมพวเตอร์ฯ-ฉบับท-2-พ-ศ-2560.aspx>

Ahmed, S., Hinkelmann, K., & Corradini, F. (2022). Development of Fake News Model using Machine Learning through Natural Language Processing. *International Scholarly and*





- Scientific Research & Innovation*. 14(12), 454-460.
<https://doi.org/10.48550/arXiv.2201.07489>
- Andrea Carson, K.F. (2019). *The real news on 'fake news': politicians use it to discredit media, and journalists need to fight back*. Retrieved from
<https://theconversation.com/the-real-news-on-fake-news-politicians-use-it-to-discredit-media-and-journalists-need-to-fight-back-123907>
- bbc. (2017). *A brief history of fake news*. Retrieved from
<https://www.bbc.co.uk/bitesize/articles/zwcgn9q>
- Briner, R.B., & Denyer, D. (2012). Systematic review and evidence synthesis as a practice and scholarship tool. In D. M. Rousseau, *The Oxford handbook of evidence-based management* (pp. 112–129). Oxford University Press.
- Brown, T.B., et al. (2020). *Language Models are Few-Shot Learners*. CS.CL. doi:
<https://doi.org/10.48550/arXiv.2005.14165>
- Chen, B. (2020). *What is One-Hot Encoding and how to use Pandas get_dummies function*. Retrieved from: <https://towardsdatascience.com/what-is-one-hot-encoding-and-how-to-use-pandas-get-dummies-function-922eb9bd4970>
- Cueva, E., Ee, G., Iyer, A., Pereira, A.S., Roseman, A.S., & Martinez, D. (2020). Detecting Fake News on Twitter Using Machine Learning Models. *2020 IEEE MIT Undergraduate Research Technology Conference (URTC)*, 1-5. DOI:10.1109/urtc51696.2020.9668872
- Deepak, S., & Chitturi, B. (2020). Deep neural approach to Fake-News identification. *Procedia Computer Science*, 167, 2236-2243. DOI:10.1016/j.procs.2020.03.276
- Derakhshan, C. W. a. H. (2018). Thinking about 'information disorder': formats of misinformation, disinformation, and mal-information. *Journalism, 'Fake News' & Disinformation. UNESCO. 2018*. Retrieved from:
https://en.unesco.org/sites/default/files/f._jfn_d_handbook_module_2.pdf
- Devarajan, G.G., Nagarajan, S.M., Amanullah, S.I., Mary, S.A.S.A., & Bashir, A.K. (2023). AI-Assisted Deep NLP-Based Approach for Prediction of Fake News from Social Media Users. *IEEE Transactions on Computational Social Systems*, 1-11.
doi:10.1109/TCSS.2023.3259480
- Garg, S., & Kumar Sharma, D. (2022). Linguistic features-based framework for automatic fake news detection. *Computers & Industrial Engineering*, 172, 108432. doi:
<https://doi.org/10.1016/j.cie.2022.108432>





- Harbert, T. (2021). *Tapping the power of unstructured data*. Retrieved from <https://mitsloan.mit.edu/ideas-made-to-matter/tapping-power-unstructured-data>
- Jadhav, S.S., & Thepade, S.D. (2019). Fake News Identification and Classification Using DSSM and Improved Recurrent Neural Network Classifier. *Applied Artificial Intelligence*, 33(12), 1058-1068. doi:10.1080/08839514.2019.1661579
- Jain, A., Shakya, A., Khatter, H., & Gupta, A. K. (2019). A smart System for Fake News Detection Using Machine Learning. *Paper presented at the 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*. DOI:10.1109/ICICT46931.2019.8977659
- Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet – A deep convolutional neural network for fake news detection. *Cognitive Systems Research*, 61, 32-44. doi: <https://doi.org/10.1016/j.cogsys.2019.12.005>
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Group, P.-P. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4 (1), 1. <https://doi.org/10.1186/2046-4053-4-1>
- Mookdarsanit, P., & Mookdarsanit, L. (2021). The COVID-19 fake news detection in Thai social texts. *Bulletin of Electrical Engineering and Informatics*.10(2), 988-998. DOI:10.11591/eei.v10i2.2745
- Nichols, J. (2020). Freedom of the Press, Fake News, and Disinformation. Retrieved from: <https://www.uua.org/international/blog/freedom-press-fake-news-disinformation>
- Panyachot. (2021). *Breaking news in the digital age*. Retrieved from <https://www.depa.or.th/th/article-view/News-consumption-in-the-digital-age>
- Prasertsom, P. (2020). *Text Extraction with Basic Language Processing Techniques: TF-IDF, Part 1*. Retrieved from <https://bigdata.go.th/big-data-101/tf-idf-1/>
- Prasertsom, P. (2021). *Finding Semantic Placeholders of Text: Word2Vec Word Embedding, Part I*. Retrieved from <https://bigdata.go.th/big-data-101/word2vec/>
- Sedik, A., Abohany, A.A., Sallam, K.M., Munasinghe, K., & Medhat, T. (2022). Deep fake news detection system based on concatenated and recurrent modalities. *Expert Systems with Applications*, 208, 117953. doi: <https://doi.org/10.1016/j.eswa.2022.117953>

