**Harnessing Large Language Models for Social Media Insights in Construction Management**

Principal Investigator (PI):

* Ida Ferding
* Graduate Student in the Master of Science in Engineering, Construction Management Program
* Florida Gulf Coast University (FGCU)
* U.A. Whitaker College of Engineering
* Fort Myers, FL, USA
* Email: [ikferding6920@eagle.fgcu.edu](mailto:ikferding6920@eagle.fgcu.edu)
* Phone: (305) 898-1270

Advisors:

* Ahmed S. Elshall, Ph.D.
  + Assistant Professor
  + Department of Bioengineering, Civil Engineering, and Environmental Engineering
  + Joint Appointment with The Water School | Florida Gulf Coast University
  + Fort Myers, FL, USA
  + Email: [aelshall@fgcu.edu](mailto:aelshall@fgcu.edu)
  + Phone: (239) 590-7591
* Anh D. Chau, Ph.D.
  + Assistant Professor
  + Stock Development Department of Construction Management
  + U.A. Whitaker College of Engineering | Florida Gulf Coast University
  + Fort Myers, FL, USA
  + Email: achau@fgcu.edu
  + Phone: (239) 590-7846

Project Details:

* Proposed Start Date: January 2025
* Proposed End Date: May 2025
* Budget: Self-funded as part of the academic requirements for the master’s in engineering

program in construction.

**PROJECT SUMMARY**

Background. The construction industry is at the forefront of a transformative era, driven by advancements in digital tools and data analytics. Social media platforms such as Reddit and YouTube have become valuable sources of real-time, unfiltered insights into industry practices, challenges, and innovations. Despite their potential, the vast and unstructured nature of these platforms makes them challenging to analyze systematically. This research addresses this gap by employing artificial intelligence (AI) to extract and synthesize thematic insights from social media data, providing actionable knowledge for stakeholders in the construction industry.

Objectives. This project aims to harness AI-driven methods to analyze construction-related discussions and multimedia content from Reddit and YouTube. The primary objectives are to identify recurring themes, emerging trends, and patterns within these platforms to provide insights that inform decision-making and strategic planning and to develop scalable methodologies for analyzing unstructured social media data. By doing so, the research seeks to bridge knowledge gaps and support innovation in workforce development, safety practices, and technology integration.

Methods. The research adopts a robust methodological framework combining Python-based programming with AI tools such as natural language processing and generative algorithms. Reddit and YouTube data from 2019 to 2025 will be collected and preprocessed to ensure relevance and integrity. The data will undergo thematic and sentiment analyses to uncover patterns and trends. Visualization techniques, including thematic charts and trend graphs, will be employed to effectively communicate findings. This interdisciplinary approach ensures a comprehensive understanding of industry dynamics while advancing the utility of AI in construction management research.

Expected Results. The study anticipates uncovering key themes, such as career development, safety concerns, technical problem-solving, and the adoption of emerging technologies. Insights into engagement dynamics and sentiment patterns will provide a deeper understanding of how professionals use social media for knowledge sharing and problem-solving. The findings are expected to support decision-making, predict industry trends, and identify areas for targeted training and education. The study will also establish a replicable framework for analyzing unstructured data, setting a foundation for future research.

Outcomes. This research will enhance the understanding of how social media platforms contribute to the construction industry’s knowledge system. It will provide actionable insights for improving workforce training, fostering innovation, and addressing critical challenges such as labor shortages and workplace safety. By integrating multimodal data from Reddit and YouTube, the study offers a more nuanced perspective on industry trends, supporting both academic exploration and practical applications.

**INTRODUCTION**

The construction industry is undergoing a transformative era fueled by digital innovation and the advancement of data-driven decision-making tools. Among the upcoming technologies reshaping the field, artificial intelligence (AI) stands out as a powerful instrument for analyzing vast and unstructured data sources (Abioye et al., 2021). Social media platforms, such as Reddit and YouTube, provide dynamic, real-time collections of industry-related discussions, insights, and challenges, creating untapped opportunities for leveraging AI in construction management research (Azhar & Abeln, 2014; Iqbal et al., 2021).

The construction industry faces a range of persistent challenges, including labor shortages, the integration of emerging technologies, and growing demands for workplace safety. To address these issues effectively, it is essential to harness new sources of data that provide real-time, diverse perspectives. Social media has emerged as a powerful tool for capturing informal exchanges that reflect the real-world challenges and innovations within the industry. However, the vast and unstructured nature of social media data often makes it difficult to process and analyze efficiently.

Reddit is a social media platform organized into topic-specific communities, or subreddits, where users share content and engage in discussions (Reddit, 2024). The subreddit r/Construction, for example, is a vibrant hub for construction professionals, covering a variety of topics such as safety protocols, career advice, and technical problem-solving. Reddit’s appeal lies in its ability to capture grassroots, unfiltered discussions that reflect the day-to-day realities and priorities of industry practitioners. By analyzing data from Reddit, this research aims to uncover evolving trends and shared knowledge that traditional studies might overlook, offering a deeper understanding of the industry's dynamic challenges and innovations.

YouTube is a global video-sharing platform that has become an essential resource for education and professional development (Duffy, 2008). In the construction industry, YouTube hosts a wealth of tutorials, equipment demonstrations, and expert discussions. These videos often provide visual, experiential insights that complement the textual exchanges found on other platforms (Nasir & Bargstädt, 2017). By incorporating YouTube data, this research captures a multimodal perspective on industry practices and knowledge dissemination, enabling a richer, more comprehensive analysis of construction trends and innovations.

AI refers to computational techniques designed to replicate human intelligence, including natural language processing (NLP), machine learning (ML), and generative algorithms. AI is particularly suited for processing unstructured data, such as social media discussions, where patterns and themes can be difficult to identify manually (Yalamati & Batchu, 2024; Uddin, 2024). This research employs AI to automate the extraction and synthesis of thematic insights from large datasets, overcoming the limitations of traditional qualitative methods (Deshpande & Kumar, 2018). AI’s scalability and efficiency allow for the identification of trends and gaps in industry knowledge, providing practical insights for professionals, educators, and policymakers.

Social media platforms like Reddit and YouTube provide rich, unstructured datasets that offer valuable insights into the construction industry's challenges, innovations, and best practices (Tang et al., 2017). However, traditional methods of analyzing social media often fall short due to the vastness and complexity of the data (Zeng et al., 2010). This research addresses these challenges by leveraging AI to enhance the efficiency and depth of analysis. By integrating multimodal data from text-based discussions and video content, the study aims to develop a more comprehensive understanding of industry trends and community engagement. This expanded approach not only captures the diverse ways professionals share and consume knowledge but also facilitates more informed decision-making and strategic planning in the construction sector. Through these efforts, the study aims to contribute to both academic understanding and practical advancements, fostering innovation and collaboration in the construction industry.

This research seeks to explore the potential of integrating AI with social media analytics to address critical challenges in the construction industry (Morgan, 2023). Using Python, a versatile programming language to access and analyze social media data through platform APIs, the study processes and analyzes discussions and content on platforms like Reddit and YouTube, enabling efficient extraction of thematic trends and insights. By analyzing discussions and content on platforms like Reddit and YouTube, this study aims to extract valuable insights into contemporary trends, industry concerns, and knowledge-sharing dynamics. These platforms capture a broad spectrum of professional discourse, offering unique perspectives that traditional methods, such as surveys or interviews, may fail to reveal.

**METHOD**

This study employs a comprehensive data collection and analysis framework leveraging Python programming and advanced AI techniques to examine construction-related content from Reddit and YouTube over a six-year span (2019–2025). The methodological process comprises four key stages: data collection, preprocessing, analysis, and visualization.

**Data Collection**: The first step is to capture a representative sample of construction-related discussions and multimedia content.

* Reddit Data: Reddit data from 2019 to 2025 will be gathered using Python and the Reddit API. This includes posts and comments from construction-focused subreddits such as r/Construction. The selection of Reddit is informed by its extensive and dynamic discourse among industry professionals, covering topics such as technical queries, safety practices, and career development. Data collection will focus on metadata such as post content, comment threads, engagement metrics (upvotes, downvotes), and timestamps.
* YouTube Data: Python-based web scraping techniques will be applied to extract construction-related content on YouTube. Using the keyword "Construction," this study will collect video transcripts, comments, like/dislike ratios, and metadata such as view counts, publication dates, and user interaction patterns. The diversity and depth of YouTube content will enable the examination of visual and textual insights into industry trends and public engagement.

**Data Preprocessing**: To prepare the collected datasets for analysis, preprocessing steps will include data cleaning, normalization, and anonymization. Non-relevant content, duplicate entries, and incomplete records will be filtered out to ensure data integrity. For textual data (e.g., Reddit posts, YouTube transcripts, and comments), natural language processing (NLP) techniques will standardize formatting and remove noise, such as stop words, punctuation, and irrelevant symbols.

**Data Analysis:** Analysis will be performed using a combination of custom Python scripts and the OpenAI API for thematic extraction, sentiment analysis, and summarization. The analysis will focus on identifying recurring patterns, themes, and trends across both platforms.

* **Thematic Analysis**: The OpenAI API will be used to categorize discussions into thematic groups, such as career development, safety, technical problem-solving, and equipment recommendations. Patterns of user engagement will be analyzed to uncover dominant topics and emerging concerns.
* **Sentiment Analysis**: Sentiment analysis tools will evaluate the emotional tone of discussions, distinguishing between positive, neutral, and negative sentiments. This will provide insights into community attitudes toward various industry challenges and innovations.

**Visualization:** Finally, the processed and analyzed data will be visualized using Python libraries such as Matplotlib and Seaborn. Visual representations will include thematic distribution charts, sentiment heatmaps, and temporal trends. The visualizations aim to communicate complex data insights effectively, making them accessible for stakeholders in academia and industry.



**Figure 1. Visual Research Methods Section**

**EXPECTED RESULTS**

Outputs. **Research Findings:** The study is expected to yield significant findings that capture both the current state and future trajectory of the construction industry. Through thematic analysis of social media data, the research will identify recurring topics, including career development, safety concerns, technical problem-solving, and technology adoption. It will provide insights into how professionals engage with these platforms to share knowledge and address common challenges. Additionally, the research will uncover trends in industry priorities and sentiment, revealing how attitudes toward workplace practices and emerging technologies are evolving over time. These findings will contribute valuable context to ongoing discussions about the industry's direction and the role of digital platforms in shaping it.

Outcomes. **Enhanced Understanding:** This research aims to deepen understanding of the construction industry’s dynamics in several key areas. It will illuminate how professionals leverage social media platforms like Reddit and YouTube for informal education, networking, and problem-solving. The study will also offer a clearer picture of emerging trends, such as the digital transformation of industry practices and the growing emphasis on safety culture. By providing data-driven insights, the research will support stakeholders in making informed decisions regarding workforce planning, safety initiatives, and the adoption of innovative technologies. Ultimately, it will contribute to a more nuanced understanding of how the industry operates in a digitally connected world.

Impacts. The broader impacts of this research will extend to both academia and industry. By establishing a novel AI-driven methodology for analyzing social media data, the study will advance academic knowledge and set a precedent for future research in construction management. Practical industry applications will include actionable insights to address labor shortages, safety concerns, and skill development, ensuring the findings are relevant to professionals and policymakers alike. The research will also encourage the integration of AI and social media analytics as essential tools for industry analysis and strategic planning. Furthermore, the study’s contributions to training and education will enhance professional development programs, providing valuable resources to improve knowledge dissemination and support the industry’s ongoing evolution.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1/6 | 1/13 | 1/20 | 1/27 | 2/3 | 2/10 | 2/17 | 2/24 | 3/3 | 3/10 | 3/17 | 3/24 | 3/31 | 4/7 | 4/14 | 4/21 |
| Prepare Research Proposal | X | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Finalize Research Proposal |  |  | X | X |  |  |  |  |  |  |  |  |  |  |  |  |
| Data Collection Setup |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |
| Data Collection |  |  |  |  | X | X |  |  |  |  |  |  |  |  |  |  |
| Thematic Analysis |  |  |  |  |  |  | X | X |  |  |  |  |  |  |  |  |
| Data Synthesis and Insight |  |  |  |  |  |  |  | X | X |  |  |  |  |  |  |  |
| Analyze Data |  |  |  |  |  |  |  |  | X | X |  |  |  |  |  |  |
| Prepare Research Paper |  |  |  |  |  |  |  |  |  | X | X | X | X |  |  |  |
| Prepare Research Presentation |  |  |  |  |  |  |  |  |  |  |  | X | X | X |  |  |
| Finalize Research Paper |  |  |  |  |  |  |  |  |  |  |  |  |  | X | X | X |
| Present Research Paper |  |  |  |  |  |  |  |  |  |  |  |  |  | X | X | X |

**TIMELINE**

**Figure 2: Research Project Activities Schedule**

**REFERENCES**

Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Davila Delgado, J. M., Bilal, M., Akinade, O. O., & Ahmed, A. (2021). Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *Journal of Building Engineering*, *44*, 103299. https://doi.org/10.1016/j.jobe.2021.103299

Azhar, S., & Abeln, J. M. (2014). Investigating Social Media Applications for the Construction Industry. *Procedia Engineering*, *85*, 42–51. https://doi.org/10.1016/j.proeng.2014.10.527

Deshpande, A., & Kumar, M. (2018). *Artificial Intelligence for Big Data: Complete guide to automating Big Data solutions using Artificial Intelligence techniques*. Packt Publishing Ltd.

Duffy, P. (2008). Using youtube: Strategies for using new media in teaching and learning. In *Enhancing Learning Through Technology* (pp. 31–43). WORLD SCIENTIFIC. https://doi.org/10.1142/9789812799456\_0003

Iqbal, T., Khan, M., Taveter, K., & Seyff, N. (2021). Mining Reddit as a New Source for Software Requirements. *2021 IEEE 29th International Requirements Engineering Conference (RE)*, 128–138. https://doi.org/10.1109/RE51729.2021.00019

Morgan, D. L. (2023). Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT. *International Journal of Qualitative Methods*, *22*, 16094069231211248. https://doi.org/10.1177/16094069231211248

Nasir, A. R., & Bargstädt, H.-J. (2017). An Approach to Develop Video Tutorials for Construction Tasks. *Procedia Engineering*, *196*, 1088–1097. https://doi.org/10.1016/j.proeng.2017.08.066

Reddit. (2024). *Dive into anything*. https://redditinc.com/

Tang, L., Zhang, Y., Dai, F., Yoon, Y., Song, Y., & Sharma, R. S. (2017). Social Media Data Analytics for the U.S. Construction Industry: Preliminary Study on Twitter. *Journal of Management in Engineering*, *33*(6), 04017038. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000554

Uddin, M. K. S. (2024). *A Review of Utilizing Natural Language Processing and Ai for Advanced Data Visualization in Real-time Analytics* (SSRN Scholarly Paper No. 4993556). Social Science Research Network. https://doi.org/10.62304/ijmisds.v1i04.185

Yalamati, S., & Batchu, R. K. (2024). Smart Data Processing: Unleashing the Power of AI and ML. In *Practical Applications of Data Processing, Algorithms, and Modeling* (pp. 205–221). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-2909-2.ch016

Zeng, D., Chen, H., Lusch, R., & Li, S.-H. (2010). Social Media Analytics and Intelligence. *IEEE Intelligent Systems*, *25*(6), 13–16. IEEE Intelligent Systems. https://doi.org/10.1109/MIS.2010.151