Literature Review

AI

The graduation project by Nguyễn Thị Hồng Hạnh, titled *"Building an Intelligent Cross-Platform Chat Bot System,"* focuses on designing and implementing a scalable, AI-powered chatbot capable of seamless operation across web, mobile, and messaging platforms. Leveraging advanced technologies like Natural Language Processing (NLP), Artificial Intelligence (AI), and tools such as React.js, Flask, and MongoDB, the system aims to deliver human-like responses, ensure cross-platform compatibility, and enhance user engagement. The project integrates models for intent detection, entity recognition, and response generation while incorporating image generation capabilities using state-of-the-art frameworks like Stable Diffusion. Key achievements include a modular architecture for scalability, multilingual support, and integration with external APIs for extended functionality. Despite its success in addressing challenges like platform dependency and limited automation in traditional chatbots, limitations such as contextual retention, performance under heavy loads, and domain-specific accuracy remain. Future enhancements proposed include advanced context management, bias mitigation, offline capabilities, and improved scalability to make the system more robust and adaptable for diverse applications such as customer service, education, healthcare, and e-commerce.

Nguyễn Thị Hồng Hạnh. \*"Building an Intelligent Cross-Platform Chat Bot System."\* Graduation Project, Vietnam-Korea University of Information and Communication Technology, supervised by Dr. Tran Van Dai, December 2024.

The paper, *"An Automated Apply Platform for Employment Junction,"* presents an advanced job portal system designed to streamline the recruitment process for job seekers and employers. It introduces features like auto-apply functionality, AI-driven job matching, customizable filters, and application tracking to enhance user experience and efficiency. The system leverages automation to reduce manual effort, improve job recommendations, and ensure seamless interaction between users. Ethical considerations, privacy measures, and mobile accessibility are also emphasized.

Dehadi, Gauri, Rishabh Choudhar, Saif Ali Sayyed, Priyadarshan Khavtode, and Dashrath Wagmare. *"An Automated Apply Platform for Employment Junction."* Graduation Project, JSPM's Bhivarabai Sawant Institute of Technology and Research, Wagholi, Maharashtra, India. February 2025.

The paper *"Artificial Intelligence Enabled, Social Media Leveraging Job Matching System for Employers and Applicants"* explores a machine learning-based system that integrates LinkedIn and Twitter data to assess job candidates' technical and emotional intelligence. By analyzing LinkedIn profiles for technical skills and Twitter posts for emotional aptitude, the system provides dual-perspective evaluations for employers and self-assessment tools for candidates. It uses NLP, regression models, and sentiment analysis to generate employability scores, emotional quotient indicators, and ranked candidate recommendations. The system aims to enhance hiring efficiency while addressing ethical concerns like bias and data privacy.

Pendyala, Vishnu, Nishtha Atrey, Tina Aggarwal, and Saumya Goyal. "Artificial Intelligence Enabled, Social Media Leveraging Job Matching System for Employers and Applicants." In *Proceedings of the 2022 International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC 2022)*. IEEE, 2022. <https://doi.org/10.1109/ICMACC54824.2022.10093323>

The working paper *"Artificial Intelligence and Job Automation: An EU Analysis Using Online Job Vacancy Data"* by Konstantinos Pouliakas examines the impact of AI and automation on European labor markets using the Skills-OVATE database of online job advertisements. The study identifies tasks with high automation risks, such as routine manual activities, while highlighting tasks reliant on socioemotional, managerial, and problem-solving skills as less automatable. Machine learning models predict occupational automation risk with about 70% accuracy, offering insights for reskilling policies to address digital transformation challenges.

Pouliakas, Konstantinos. *Artificial Intelligence and Job Automation: An EU Analysis Using Online Job Vacancy Data.* Luxembourg: Publications Office of the European Union, 2021. Cedefop Working Paper No. 6. <https://doi.org/10.2801/305373>

The paper *"Artificial Intelligence for Job Seeking: How to Enhance Labor Intermediation in Public Employment Services"* by Manuel Urquidi and Gloria Ortega examines how AI can improve Public Employment Services (PES) by addressing inefficiencies in matching labor supply and demand. It highlights AI's potential in profiling candidates, identifying skill gaps, and providing tailored recommendations for job seekers, including vulnerable groups. The study also discusses challenges such as data privacy, algorithmic bias, and digital accessibility, proposing strategies to mitigate these risks while enhancing PES efficiency and inclusivity.

Urquidi, Manuel, and Gloria Ortega. *Artificial Intelligence for Job Seeking: How to Enhance Labor Intermediation in Public Employment Services.* IDB Technical Note No. 1996. Washington, DC: Inter-American Development Bank, 2020.

The research paper *"Tawzef: Improving Recruitment Portals Performance via AI Technology"* by Magdy Elhennawy and Layla Reda proposes an AI-driven framework to enhance online job portals. It introduces advanced resume recognition using Natural Language Processing (NLP) and heuristic approaches, combined with matching algorithms like skills count and cosine similarity. The study demonstrates that heuristic recognition paired with skills count matching significantly reduces processing time, improving recruitment efficiency. The findings emphasize AI's role in automating tedious tasks, minimizing errors, and optimizing the job-matching process.

Elhennawy, Magdy, and Layla Reda. *Tawzef: Improving Recruitment Portals Performance via AI Technology.* Higher Institute of Computers and Information Technology, El Shorouk Academy, Cairo, Egypt, 2025.

The paper *"Ensuring Sustainable Growth Based on the Artificial Intelligence Analysis and Forecast of In-Demand Skills"* by Alena Vankevich and Iryna Kalinouskaya proposes an AI-driven framework to analyze and forecast labor market skills using data scraped from online job portals. It identifies gaps between employer demands and candidate skills, emphasizing the need for AI-based classification and comparison of competencies. This approach aims to improve education policies, labor market forecasting, and workforce training for sustainable growth.

Vankevich, Alena, and Iryna Kalinouskaya. *"Ensuring Sustainable Growth Based on the Artificial Intelligence Analysis and Forecast of In-Demand Skills."* E3S Web of Conferences 208 (2020): 03060. <https://doi.org/10.1051/e3sconf/202020803060>

The paper, *Building a Job Portal Using Web Scraping* by Vishnu Priya N. et al., presents a web application that simplifies job searching by leveraging web scraping techniques. The platform automates data extraction from job portals, consolidates listings, and provides personalized recommendations using a job search and recommendation engine. Key features include user-friendly registration, email notifications, and filtering options. Challenges like ethical considerations in web scraping and technical hurdles (e.g., email notifications) were addressed using tools like Python's Beautiful Soup and SMTP. The system enhances efficiency for job seekers by reducing manual effort while ensuring ethical data use1.

Vishnu Priya N., Nehaa Shri A., Mary Valentina Janet A., and Mubeen Fathima G. "Building a Job Portal Using Web Scraping." *International Journal of Progressive Research in Engineering Management and Science (IJPREMS)* 3, no. 5 (May 2023): 633-637. doi:10.58257/IJPREMS31247.

The paper *"A Comprehensive Framework for Online Job Portals for Job Recommendation Strategies Using Machine Learning Techniques"* by Kamal Upreti et al. proposes an AI-driven job recommendation system leveraging machine learning methods like Random Forest, K-Nearest Neighbor (KNN), and Support Vector Machines (SVM). By analyzing user resumes and job descriptions, the system employs content-based filtering to provide personalized job suggestions. Random Forest emerged as the most effective algorithm, offering high accuracy and efficiency in matching candidates with suitable opportunities.

Upreti, Kamal, Shikha Mittal, Prakash Divakaran, Prashant Vats, Manpreet Bajwa, and Sandeep Singh. *"A Comprehensive Framework for Online Job Portals for Job Recommendation Strategies Using Machine Learning Techniques."* In *ICT Infrastructure and Computing: Proceedings of ICT4SD 2022.*

The paper, *Implementation of an Intelligent Online Job Portal Using Machine Learning Algorithms* by Zarrin Tasnim et al., proposes a system integrating three phases: extracting job circulars using web crawlers, clustering similar job searches via K-means, and sending email notifications based on decision tree algorithms. The system automates data collection, analyzes job market demand, and matches job seekers to relevant opportunities. It enhances efficiency in job searching and skill development while aiding HR departments in hiring decisions.

Tasnim, Zarrin, F. M. Javed Mehedi Shamrat, Shaikh Muhammad Allayear, Khobayeb Ahmed, and Naimul Islam Nobel. "Implementation of an Intelligent Online Job Portal Using Machine Learning Algorithms." *International Journal of Scientific & Technology Research* 9, no. 2 (2020): 2450-2557.

The paper *"Emotion AI: Integrating Emotional Intelligence with Artificial Intelligence in the Digital Workplace"* by Simran Kaur and Richa Sharma explores the integration of Emotional Intelligence (EI) with Artificial Intelligence (AI) to improve workplace processes. It highlights how AI technologies, such as emotion recognition and affective computing, can analyze non-verbal cues like facial expressions and voice to simulate human emotions. Applications in recruitment, onboarding, training, healthcare, and retail are discussed, emphasizing AI's role as a support system for EI rather than a replacement.

Kaur, Simran, and Richa Sharma. *"Emotion AI: Integrating Emotional Intelligence with Artificial Intelligence in the Digital Workplace."* In *Innovations in Information and Communication Technologies (IICT-2020)*. Amity University, Noida, Uttar Pradesh, India.

The paper *"Job Portal End Detection of Fake Job Posting Using Machine Learning"* by Prem Anand and Dr. Vishnu Sharma presents a machine learning-based model to detect fraudulent job postings on online job portals. Using a Kaggle dataset, the study employed data preprocessing, feature selection, and a decision tree algorithm to classify job postings as fake or genuine. Key features like company CIN number, official email ID, and security deposit were critical for prediction. The model achieved an accuracy of 97.5%, demonstrating its potential to enhance trust and security in online recruitment systems.

Citation (Chicago Style)

Anand, Prem, and Vishnu Sharma. *"Job Portal End Detection of Fake Job Posting Using Machine Learning."* M.Tech Dissertation, Jaipur National University, Jaipur, 2025.

The paper, *Job Portal Resume Evaluation System Using Text Mining and Natural Language Processing* by Jaichandran R. et al., introduces a job portal that streamlines recruitment via text mining and NLP. It includes features like resume de-duplication (using Proactive Replica Checking), resume scoring based on education and skills, and collaborative filtering for job recommendations. Employers can upload study materials, while job seekers benefit from tailored job matches. The system optimizes storage costs and simplifies hiring processes.**Citation (Chicago Style):**  
Jaichandran, R., S. Leelavathy, Kanaga Suba Raja, Pranav Kumar, Sailendra Kumar Majhi, and Vikas Singh Thakur. "Job Portal Resume Evaluation System Using Text Mining and Natural Language Processing." *European Journal of Molecular & Clinical Medicine* 7, no. 4 (2020): 2450-2557.

The paper *"JRC: A Job Post and Resume Classification System for Online Recruitment"* by Abeer Zaroor, Mohammed Maree, and Muath Sabha presents a hybrid system designed to enhance e-recruitment by classifying resumes and job posts into occupational categories using an integrated knowledge base. The system employs Natural Language Processing (NLP) for segmentation, conceptual classification, and semantic matching to improve precision and reduce runtime complexity. Experimental results demonstrate the system's efficiency in achieving accurate matches between resumes and job postings compared to conventional methods.

Citation (Chicago Style)

Zaroor, Abeer, Mohammed Maree, and Muath Sabha. *"JRC: A Job Post and Resume Classification System for Online Recruitment."* Conference Paper, Proceedings of the 2017 IEEE International Conference on Tools with Artificial Intelligence (ICTAI), November 2017. <https://doi.org/10.1109/ICTAI.2017.00123>

The paper *"Artificial Intelligence: A Technological Prototype in Recruitment"* by R. Vedapradha, Ravi Hariharan, and Rajan Shivakami evaluates the adaptability of AI in recruitment and its impact on employee performance. Using a sample of 440 HR professionals from Bangalore, the study employs multiple linear regression and ANOVA to analyze variables like productivity, training, automation, reliability, and gamification. Findings indicate that AI significantly enhances productivity, training efficiency, and automation while reducing recruitment costs. However, gamification showed a limited impact. AI's integration in recruitment automates repetitive tasks, improves decision-making, and enhances candidate engagement.

Citation (Chicago Style)

Vedapradha, R., Ravi Hariharan, and Rajan Shivakami. "Artificial Intelligence: A Technological Prototype in Recruitment." *Journal of Service Science and Management* 12, no. 3 (2019): 382–390. <https://doi.org/10.4236/jssm.2019.123026>.

The paper, *Data Mining of Job Requirements in Online Job Advertisements Using Machine Learning and SDCA Logistic Regression* by Bogdan Walek and Ondrej Pektor, proposes a system to extract job requirements from online job advertisements. It integrates a data mining module, an SDCA logistic regression-based machine learning module, and postprocessing techniques to enhance accuracy. Tested on 9,971 IT job advertisements, the system achieved an 80% success rate in identifying job requirements. It also highlights frequent IT skills and compares results with the Open Skills database.**Citation (Chicago Style):**  
Walek, Bogdan, and Ondrej Pektor. "Data Mining of Job Requirements in Online Job Advertisements Using Machine Learning and SDCA Logistic Regression." *Mathematics* 9, no. 19 (2021): 2475. <https://doi.org/10.3390/math9192475>.

The paper, *Online Job Portal* by Md. Ahmed et al., addresses the growing issue of fraudulent job postings on online platforms. It proposes a machine learning-based system employing algorithms like KNN, decision trees, SVM, Naive Bayes, random forests, and deep neural networks to classify job postings as genuine or deceptive. The study highlights the superiority of ensemble classifiers over single classifiers for scam detection. By leveraging NLP and classification techniques, the system aims to enhance job portal security and protect job seekers.**Citation (Chicago Style):**  
Ahmed, Md., Bolem Lakshmi Meghana, Naraharasetti Sasi Durga, Matta Naga Sri, and Shaik Shabuddin. "Online Job Portal." *International Journal of Engineering Research and Science & Technology* 20, no. 2 (April 2024): 1-10.

The paper *"Use of Online Job Portal Data in Research and in Practice: A Review"* by Merl Chandana and Vihanga Jayawickrama explores the potential of online job portals (OJPs) as a rich data source for labor market analysis. It highlights applications such as monitoring skill demand, identifying emerging trends, and testing sociological theories. Despite their advantages, challenges like representativity, data quality, and privacy concerns are discussed, along with solutions like statistical adjustments and complementary data sources. The review emphasizes the need for standardized methodologies to enhance the reliability of OJP-based research.

Citation (Chicago Style)

Chandana, Merl, and Vihanga Jayawickrama. *"Use of Online Job Portal Data in Research and in Practice: A Review."* LIRNEasia, July 2022.

The paper *"Artificial Intelligence Chatbots are New Recruiters"* by Nishad Nawaz and Anjali Mary Gomes examines the impact of AI chatbots on the recruitment process. It highlights how chatbots enhance candidate engagement, streamline tasks like pre-screening, scheduling, and onboarding, and improve recruitment efficiency. The study emphasizes chatbots' ability to address complex recruitment challenges while reducing human workload. It also identifies opportunities for future research in chatbot-driven recruitment strategies and their broader implications for HR management.

Citation (Chicago Style)

Nawaz, Nishad, and Anjali Mary Gomes. *"Artificial Intelligence Chatbots are New Recruiters."* Department of Business Management, College of Business Administration, Kingdom University, Riffa, Bahrain, 2025.

The paper, *The Use of Artificial Intelligence in Job Seeking and Competence Development* by Markko Liutkevičius and Sadok Ben Yahia, explores AI applications in job and training recommendation systems. It highlights AI techniques such as Machine Learning, Neural Networks, and Deep Learning for personalized job recommendations and training suggestions. The study proposes a citizen-centered AI-enabled service architecture for the European labor market, particularly targeting individuals over 50. It also emphasizes integrating AI into platforms like Silver Hub to enhance e-governance and career services.**Citation (Chicago Style):**  
Liutkevičius, Markko, and Sadok Ben Yahia. "The Use of Artificial Intelligence in Job Seeking and Competence Development." *Proceedings of the AHFE International Conference on Human Factors in Software and Systems Engineering*, January 2022. <https://doi.org/10.54941/ahfe1002260>.

WEB SCRAPING

1. The chapter *"Web Scraping"* by Bo Zhao provides an overview of web scraping as a technique for extracting data from the web and organizing it into structured formats for analysis. It outlines the two main steps: acquiring web resources via HTTP requests and extracting information using tools like Beautiful Soup, Pyquery, and Scrapy. The chapter discusses various applications of web scraping, such as price monitoring, sentiment analysis, and search engine indexing, while addressing legal and ethical concerns like copyright infringement, terms of service violations, and server overloading. It also highlights anti-scraping measures like HTML fingerprinting, CAPTCHA challenges, and IP reputation checks.

Citation (Chicago Style)

Zhao, Bo. "Web Scraping." In *Encyclopedia of GIS*, edited by Shashi Shekhar, Hui Xiong, and Xun Zhou, 1–9. Cham: Springer, 2017. <https://doi.org/10.1007/978-3-319-32001-4_483-1>.

1. The paper, *The Use of Artificial Intelligence in Job Seeking and Competence Development* by Markko Liutkevičius and Sadok Ben Yahia, explores AI applications in job and training recommendation systems. It highlights AI techniques such as Machine Learning, Neural Networks, and Deep Learning for personalized job recommendations and training suggestions. The study proposes a citizen-centered AI-enabled service architecture for the European labor market, particularly targeting individuals over 50. It also emphasizes integrating AI into platforms like Silver Hub to enhance e-governance and career services.**Citation (Chicago Style):**  
   Liutkevičius, Markko, and Sadok Ben Yahia. "The Use of Artificial Intelligence in Job Seeking and Competence Development." *Proceedings of the AHFE International Conference on Human Factors in Software and Systems Engineering*, January 2022. <https://doi.org/10.54941/ahfe1002260>.
2. The paper, *Design and Visualization of Python Web Scraping Based on Third-Party Libraries and Selenium Tools* by Shujun Yuan, explores Python-based web scraping for analyzing data from Chinese movie websites. Using tools like Selenium and libraries such as Matplotlib, the study captures, cleans, and visualizes data on movie genres and ratings. It highlights trends like the popularity of drama films and the overlooked status of adventure films. The research also integrates a GUI for user interaction and emphasizes Python's efficiency in data analysis and visualization.**Citation (Chicago Style):**  
   Yuan, Shujun. "Design and Visualization of Python Web Scraping Based on Third-Party Libraries and Selenium Tools." *School of Information Engineering, Hubei University of Economics*, 2025.
3. The paper *"Web Scraping Job Portals"* by Ashutosh Kumar, Kinshuk Chauhan, and Jaspreet Kaur Grewal explores the use of web scraping technology to create a job portal that aggregates employment opportunities from various sources. It employs tools like Beautiful Soup, Selenium, and Scrapy to extract job data and store it in a MongoDB database. The platform offers features such as personalized job recommendations, email notifications, and advanced search filters to enhance user experience. The study emphasizes data privacy and security while proposing measures like encryption and access control to ensure user trust.

Citation (Chicago Style)

Kumar, Ashutosh, Kinshuk Chauhan, and Jaspreet Kaur Grewal. *"Web Scraping Job Portals."* Chandigarh University, Mohali, Punjab, India, 2025.

1. The paper *"Smart Job Search Engine Using Web Scraping"* by Saumya Sinha, Rashee Saxena, and Hitesh Kr Garg introduces a Python-based web application that utilizes web scraping and crawling techniques to fetch job postings from platforms like Naukri.com and Indeed.com. Unlike traditional systems, it does not store scraped data locally but dynamically retrieves and displays results based on user inputs such as job preferences and location. The system achieves a 95% accuracy in information retrieval, providing an efficient and user-friendly solution for job seekers, especially during the COVID-19 pandemic.

Citation (Chicago Style)

Sinha, Saumya, Rashee Saxena, and Hitesh Kr Garg. *"Smart Job Search Engine Using Web Scraping."* International Journal of Advances in Engineering and Management 3, no. 6 (June 2021): 586–591.

1. The paper, *Design and Visualization of Python Web Scraping Based on Third-Party Libraries and Selenium Tools* by Shujun Yuan, explores Python-based web scraping for analyzing data from Chinese movie websites. Using tools like Selenium and libraries such as Matplotlib, the study captures, cleans, and visualizes data on movie genres and ratings. It highlights trends like the popularity of drama films and the overlooked status of adventure films. The research also integrates a GUI for user interaction and emphasizes Python's efficiency in data analysis and visualization.**Citation (Chicago Style):**  
   Yuan, Shujun. "Design and Visualization of Python Web Scraping Based on Third-Party Libraries and Selenium Tools." *School of Information Engineering, Hubei University of Economics*, 2025.
2. The paper *"Web Scraping (IMDB) Using Python"* by Narendra Kumar Rao et al. discusses the development of a Python-based web scraping system for extracting and analyzing data from the IMDB website. Using libraries like BeautifulSoup, Pandas, and Matplotlib, the system retrieves movie-related data, processes it into structured formats, and generates visualizations for insights such as popular genres and ratings over time. The study emphasizes the utility of web scraping in data-driven decision-making while addressing challenges like dynamic web structures and API limitations.

Citation (Chicago Style)

*"Web Scraping (IMDB) Using Python."* Telematique 21, no. 1 (2022): 235–247.

1. The paper, *Web Scraping Using Python* by Sanit Kumar et al., discusses the extraction of structured data from unstructured web pages using Python libraries like BeautifulSoup, Selenium, Requests, and Pandas. It explains the differences between static and dynamic web scraping, emphasizing the complexity of JavaScript-rendered pages. The study details the setup process, including virtual environments and library installation, and demonstrates how extracted data can be stored in structured formats such as CSV or JSON for further use.**Citation (Chicago Style):**  
   Kumar, Sanit, Jyoti Thakur, Deepmala Ekka, and Ishwar Sahu. "Web Scraping Using Python." *International Journal of Advances in Engineering and Management (IJAEM)* 4, no. 9 (September 2022): 235-237.
2. The paper, *An Overview of Web Scraping: Technical Aspects and Exercises* by Gustavo Pérez Molano, reviews web scraping techniques, tools, and their ethical and legal implications. It explores three methods: Python's BeautifulSoup, Octoparse, and ParseHub, comparing their effectiveness on static and dynamic websites. The study highlights the importance of understanding website structures, security measures like CAPTCHA, and ethical considerations. Results show commercial tools like Octoparse excel in ease of use, while Python offers flexibility for programmers.**Citation (Chicago Style):**  
   Pérez Molano, Gustavo. "An Overview of Web Scraping: Technical Aspects and Exercises." Electrical and Computer Engineering and Computer Science Department, Polytechnic University of Puerto Rico, 2025.
3. The paper "Performance Analysis for Web Scraping Tools: Case Studies on BeautifulSoup, Scrapy, HtmlUnit, and Jsoup" by Yılmaz Dikilitaş et al. evaluates the performance of four popular web scraping tools based on metrics like memory usage, CPU utilization, and execution time. BeautifulSoup and Jsoup excel in efficiency with low resource consumption, while Scrapy balances advanced features and moderate resource use. HtmlUnit stands out for handling dynamic content but requires higher CPU usage. The study provides insights to guide developers in selecting tools based on project-specific requirements.

Citation (Chicago Style)

Dikilitaş, Yılmaz, Çoşkun Çakal, Ahmet Can Okumuş, Halime Nur Yalçın, Emine Yıldırım, Ömer Faruk Ulusoy, Bilal Macit, Aslı Ece Kırkaya, Özkan Yalçın, Ekin Erdoğmuş, and Ahmet Sayar. "Performance Analysis for Web Scraping Tools: Case Studies on BeautifulSoup, Scrapy, HtmlUnit, and Jsoup." In Advances in Information Systems Development, edited by Springer Nature Switzerland AG, April 2024. <https://doi.org/10.1007/978-3-031-56728-5_39>.

1. The paper, Automating Hidden Gambling Detection in Websites: A BeautifulSoup Implementation by Prasert Teppap et al., proposes a system to detect hidden gambling advertisements on Thai university websites. Using Python's BeautifulSoup, the system automates web scraping, keyword detection, and reporting via LINE Notify. It achieved an accuracy of 89%, surpassing traditional methods by 53%. The study highlights the need for robust cybersecurity measures in public institutions and suggests future enhancements like predictive analytics for improved adaptability.**Citation (Chicago Style):**  
   Teppap, Prasert, Panudech Tipauksorn, Somnuek Surathong, and Prasert Luekhong. "Automating Hidden Gambling Detection in Websites: A BeautifulSoup Implementation." Proceedings of the 2024 IEEE International Conference on Computer Science and Software Engineering, June 2024. <https://doi.org/10.1109/JCSSE61278.2024.10613687>.
2. The paper "Web Crawling on News Web Page using Different Frameworks" by Harshala Bhoir and K. Jayamalini compares the performance of two Python-based web crawling frameworks, Scrapy and BeautifulSoup, for extracting data from news websites. The study highlights Scrapy's efficiency in handling large-scale asynchronous crawling and BeautifulSoup's simplicity for small-scale tasks. Experimental results show that BeautifulSoup requires less time (1.32 seconds) compared to Scrapy (1.63 seconds) for crawling the NDTV news website, emphasizing their suitability for different use cases.

Citation (Chicago Style)

Bhoir, Harshala, and K. Jayamalini. "Web Crawling on News Web Page using Different Frameworks." International Journal of Scientific Research in Computer Science, Engineering and Information Technology 7, no. 4 (July-August 2021): 513–519. https://doi.org/10.32628/CSEIT2174120.

1. The paper "Web Scraping Approaches and their Performance on Modern Websites" by Ajay Sudhir Bale et al. evaluates seven web scraping methods, including Python's requests library, Selenium (with variations like headless mode), and the undetected-chromedriver library, across 120 websites from eight categories. The study measures parameters like detection time, request count before blocking, and data extraction success. Results show that undetected-chromedriver outperformed other methods in bypassing anti-bot mechanisms, highlighting the need for improved website protection against scraping.

Citation (Chicago Style)

Bale, Ajay Sudhir, Naveen Ghorpade, Rohith S., S. Kamalesh, Rohith R., and Rohan B. S. "Web Scraping Approaches and their Performance on Modern Websites." In Proceedings of the 2022 International Conference on Electronics and Sustainable Communication Systems (ICESC), September 2022. https://doi.org/10.1109/ICESC54411.2022.9885689.

1. The paper, *Leveraging Python for Web Scraping and Data Analysis: Applications, Challenges, and Future Directions* by M. Sandeep Kumar et al., explores Python's role in web scraping and data analysis. It highlights tools like Beautiful Soup, Scrapy, Selenium, and Pandas for extracting and analyzing web data across domains such as e-commerce, finance, and healthcare. The study addresses ethical concerns, technical challenges like anti-scraping mechanisms, and emphasizes the need for responsible practices. Future advancements in AI-driven automation are also discussed as transformative for the field.**Citation (Chicago Style):**  
   Kumar, M. Sandeep, Vadla Varsha, Yeyya Prem Swarup, Sayeed Ahmad, and Kandhika Chandhu. "Leveraging Python for Web Scraping and Data Analysis: Applications, Challenges, and Future Directions." *International Journal of Data Science Research*, December 31, 2024.
2. <https://ieeexplore.ieee.org/abstract/document/10420941>
3. <https://ieeexplore.ieee.org/abstract/document/10617017>
4. <https://ieeexplore.ieee.org/abstract/document/10550225>
5. <https://ieeexplore.ieee.org/abstract/document/10823528>
6. <https://ieeexplore.ieee.org/abstract/document/10032044>
7. <https://ieeexplore.ieee.org/abstract/document/10145369>

AWS

1. The paper *"A Review on AWS - Cloud Computing Technology"* by Neha Kewate et al. provides a comprehensive overview of Amazon Web Services (AWS) as a leading cloud computing platform. It discusses AWS's core services, including S3 for storage and EC2 for virtual machines, emphasizing its scalability, cost-effectiveness, and global deployment capabilities. The study highlights AWS's robust security features, such as IAM for access control and multi-factor authentication, and compares AWS with other cloud platforms like Microsoft Azure and Google Cloud. The paper concludes by emphasizing AWS's role in addressing cloud storage and security challenges in modern enterprises.

Citation (Chicago Style)

Kewate, Neha, Amruta Raut, Mohit Dubekar, Yuvraj Raut, and Ankush Patil. *"A Review on AWS - Cloud Computing Technology."* International Journal for Research in Applied Science & Engineering Technology 13, no. 1 (2025): 1–12. <https://doi.org/10.22214/IJRASET>.

1. The paper, *An Analysis of Performance Variability in AWS Virtual Machines* by Miguel de Lima et al., investigates the performance variability of AWS Compute Optimized (C family) virtual machines across regions, instance generations, and markets. Graviton processors (c6g.12xlarge and c7g.12xlarge) demonstrated minimal variability and cost-effectiveness, outperforming Intel and AMD instances. Spot instances offered significant cost savings without performance compromises, but availability issues were noted. The study provides insights for selecting VMs based on workload requirements and introduces AWSBENCH for performance benchmarking.**Citation (Chicago Style):**  
   de Lima, Miguel, Luan Teylo, and Lúcia Drummond. "An Analysis of Performance Variability in AWS Virtual Machines." *Instituto de Computação - Universidade Federal Fluminense (UFF)* and *Centre Inria de l'université de Bordeaux - INRIA Bordeaux*, February 2025.
2. The paper "Proactive and Power Efficient Hybrid Virtual Network Embedding: An AWS Cloud Case Study" by Ikhlasse Hamzaoui et al. proposes a hybrid Virtual Network Embedding (VNE) approach to optimize resource utilization and power efficiency in AWS cloud environments. Using a Mixed Integer Linear Programming (MILP) model, the study combines proactive Virtual Node Embedding (VNoE) and multistep Virtual Link Embedding (VLiE). The approach integrates green energy prioritization through a Green-Location Aware Global Topology Ranking (GLA-GTR) algorithm. Experimental results demonstrate significant improvements in power efficiency, with reductions of up to 17.21% in power consumption compared to other methods.

Citation (Chicago Style)

Hamzaoui, Ikhlasse, Benjamin Duthil, Vincent Courboulay, and Hicham Medromi. "Proactive and Power Efficient Hybrid Virtual Network Embedding: An AWS Cloud Case Study." IEEE Access 10 (2022): 57511–57513.

1. The paper, *A Comparative Analysis of AWS Cloud-Native Application Deployment Model* by Khandakar Razoan Ahmed and Md. Motaharul Islam, evaluates Virtual Machine (VM) and Docker Container deployment models using AWS services like EC2 and ECS. Performance tests (Linear Load, Step Load, and Peak Load) conducted with Apache JMeter revealed that Docker containers outperform VMs in HTTP response and throughput, while VMs excel in error handling. Docker's lightweight architecture and scalability make it ideal for modern cloud applications.**Citation (Chicago Style):**  
   Ahmed, Khandakar Razoan, and Md. Motaharul Islam. "A Comparative Analysis of AWS Cloud-Native Application Deployment Model." In *Proceedings of the International Conference on Advanced Computing*, October 2022. <https://doi.org/10.1007/978-981-19-2445-3_29>.
2. The paper, *Comprehensive Survey of Amazon Web Services (AWS): Techniques, Tools, and Best Practices for Cloud Solutions* by Praveen Borra, provides an in-depth analysis of AWS's history, core services, architecture, and applications. It highlights AWS's scalability, cost-effectiveness, and innovation in areas like machine learning and IoT. The study explores AWS's competitive landscape, comparing it with Microsoft Azure and Google Cloud Platform, while addressing challenges like cost management and vendor lock-in. Future trends include hybrid cloud adoption and sustainability goals.**Citation (Chicago Style):**  
   Borra, Praveen. "Comprehensive Survey of Amazon Web Services (AWS): Techniques, Tools, and Best Practices for Cloud Solutions." *International Research Journal of Advanced Engineering and Science*, July 2024.
3. The paper, *AWS Cost Management and Trend Analysis* by Eetu Juvonen, focuses on automating cost reporting for Qvantel Finland Oy using AWS services. It addresses inaccuracies in client billing data by developing a Python script with the Boto3 SDK, which generates corrected cost reports and stores them in AWS S3. The project partially achieved its goals, providing accurate cost data but leaving trend analysis and full automation for future development. The script was shared via Git for team collaboration.**Citation (Chicago Style):**  
   Juvonen, Eetu. *AWS Cost Management and Trend Analysis.* Bachelor's thesis, Jyväskylä: Jamk University of Applied Sciences, May 2023.
4. The paper "AWS Data Visualization using DynamoDB and Lambda" by Ej Miguel Francisco Caliwag, Angela Caliwag, and Wansu Lim presents a temperature data sensing system leveraging AWS IoT services. Using a DHT22 sensor connected to Raspberry Pi 4B, the system collects temperature and humidity data, which is processed and stored in AWS DynamoDB via MQTT protocol. AWS Lambda triggers further integration with AWS S3 for storage, AWS QuickSight for visualization, and AWS Machine Learning for advanced analytics. The study highlights the efficiency of combining IoT devices with cloud services for real-time data processing and visualization.

Citation (Chicago Style)

Caliwag, Ej Miguel Francisco, Angela Caliwag, and Wansu Lim. "AWS Data Visualization using DynamoDB and Lambda." Conference Paper, Kumoh National Institute of Technology, July 2021.

1. The paper, *Leveraging AWS APIs for Database Scalability and Flexibility: A Case Study Approach* by Vijay Panwar, explores how AWS APIs enhance database scalability and adaptability across industries. It analyzes AWS database services like RDS, DynamoDB, and Aurora, focusing on API-driven automation for provisioning, scaling, and monitoring. Case studies from e-commerce, healthcare, and finance sectors highlight improved scalability, operational efficiency, and regulatory compliance. Challenges include API complexity and data consistency issues. Future trends emphasize AI integration and serverless architectures.**Citation (Chicago Style):**  
   Panwar, Vijay. "Leveraging AWS APIs for Database Scalability and Flexibility: A Case Study Approach." *International Journal of Engineering Applied Sciences and Technology* 8, no. 11 (March 2024): 44–52.
2. The paper "Securing DynamoDB: In-Depth Exploration of Approaches, Overcoming Challenges, and Implementing Best Practices for Robust Data Protection" by Rimsha Sajid et al. explores security measures for Amazon DynamoDB, focusing on access control, encryption, auditing, monitoring, and data integrity. It highlights threats like unauthorized access, data leakage, and injection attacks while proposing solutions such as IAM policies, fine-grained access control, encryption at rest and in transit, and real-time monitoring with AWS CloudTrail and CloudWatch. The study emphasizes balancing security with scalability to protect sensitive data in dynamic cloud environments.

Citation (Chicago Style)

Sajid, Rimsha, Gohar Mumtaz, Hijab Zehra Zaidi, and Zeeshan Mubeen. "Securing DynamoDB: In-Depth Exploration of Approaches, Overcoming Challenges, and Implementing Best Practices for Robust Data Protection." Faculty of Computer Science and Information Technology, Superior University, Lahore. Published September 1, 2024.

1. The paper, *Highly Scalable and Load Balanced Web Server on AWS Cloud* by M. Mangayarkarasi et al., demonstrates the deployment of a web server using AWS CloudFormation with YAML code to implement Infrastructure as Code (IaC). Key components include EC2 instances with Nginx, Elastic File System (EFS) for shared storage, Application Load Balancer (ALB) for traffic distribution, and Auto Scaling Groups (ASG) for dynamic scalability. The system improves performance, reduces workload, and enhances cost efficiency compared to on-premise servers.**Citation (Chicago Style):**  
   Mangayarkarasi, M., S. Tamil Selvan, R. Kuppuchamy, S. Shanthi, and S. R. Prem. "Highly Scalable and Load Balanced Web Server on AWS Cloud." *IOP Conference Series: Materials Science and Engineering* 1055, no. 1 (2021): 012113. <https://doi.org/10.1088/1757-899X/1055/1/012113>.
2. The paper "A Review on Amazon Web Service (AWS), Microsoft Azure, and Google Cloud Platform (GCP) Services" by Pooja Mittal et al. reviews the competitive landscape of cloud computing, focusing on three major providers: AWS, Microsoft Azure, and Google Cloud Platform. It examines their service offerings, pricing models, and market strategies. AWS is highlighted for its global reach and reliability, Azure for its integration with enterprise solutions, and GCP for its innovation and cost-effectiveness. The study compares their strengths and limitations, emphasizing how these platforms meet diverse customer needs while addressing cost efficiency and scalability.

Citation (Chicago Style)

Mittal, Pooja, et al. "A Review on Amazon Web Service (AWS), Microsoft Azure, and Google Cloud Platform (GCP) Services." Department of Computer Science and Engineering, SEST, Jamia Hamdard, New Delhi, 2025.

1. The paper "The Effect of Amazon Web Services (AWS) on Cloud-Computing" by Taranjot Singh examines AWS's transformative impact on cloud computing. It highlights AWS's key features, including scalability, flexibility, and security, which address challenges like data privacy, compliance, and cost-effectiveness. The study outlines AWS's service models (IaaS, PaaS, SaaS), deployment models (cloud, hybrid, on-premises), and tools like AWS Lake Formation and Migration Hub for efficient data management. It concludes that AWS's pay-as-you-go model and global reach make it a leading cloud service provider.

Citation (Chicago Style)

Singh, Taranjot. "The Effect of Amazon Web Services (AWS) on Cloud-Computing." International Journal of Engineering Research and Technology, December 2021. https://doi.org/10.356809704.

1. The paper "AWS Cloud Computing Solutions: Optimizing Implementation for Businesses" by Iqra Naseer explores strategies for optimizing Amazon Web Services (AWS) to address modern business needs. It highlights AWS's scalability, flexibility, and cost-effectiveness, showcasing its services like EC2, S3, DynamoDB, and SageMaker. Case studies of companies like Netflix and Airbnb illustrate AWS's role in fostering innovation and operational efficiency. The study also discusses challenges such as data security, migration complexity, and compliance while emphasizing future trends like automation, AI integration, and enhanced infrastructure.

Citation (Chicago Style)

Naseer, Iqra. "AWS Cloud Computing Solutions: Optimizing Implementation for Businesses." Journal of Statistics, Computing and Interdisciplinary Research 5, no. 2 (2023): 121–132.

1. The paper, *Techniques for Optimizing AWS Storage Costs and Performance* by Sri Harsha Vardhan Sanne, provides a detailed review of methods to balance AWS storage expenses while maintaining high performance. It explores strategies such as selecting appropriate storage classes (e.g., S3 Standard, Glacier), implementing lifecycle management policies, and utilizing data compression techniques like GZIP. The study also evaluates native AWS tools (e.g., CloudWatch, Trusted Advisor) and third-party solutions for monitoring and cost optimization. The findings emphasize a holistic approach to achieve efficient and cost-effective cloud storage.**Citation (Chicago Style):**  
   Sanne, Sri Harsha Vardhan. "Techniques for Optimizing AWS Storage Costs and Performance." *West Carnegie Mellon University*, February 2025.
2. The paper "Cloud Computing in Construction Industry: Use Cases, Benefits, and Challenges" by Sururah A. Bello et al. reviews the adoption and applications of cloud computing in the construction sector. It highlights cloud computing as an enabler for emerging technologies like BIM, IoT, and AR/VR, offering benefits such as cost reduction, scalability, data security, and enhanced collaboration. The study identifies challenges like latency, data privacy concerns, and poor broadband connectivity at construction sites. It also outlines future opportunities, including real-time collaboration, tighter supply chain integration, and innovative business models like "construction-as-a-service."

Citation (Chicago Style)

Bello, Sururah A., Lukumon O. Oyedele, Olugbenga O. Akinade, Muhammad Bilal, Juan Manuel Davila Delgado, Lukman A. Akanbi, Anuoluwapo O. Ajayi, and Hakeem A. Owolabi. "Cloud Computing in Construction Industry: Use Cases, Benefits, and Challenges." Automation in Construction 122 (2021): 103441. <https://doi.org/10.1016/j.autcon.2020.103441>.

1. The paper, *Identifying Challenges for Clients in Adopting Sustainable Public Cloud Computing* by Muhammad Janas Khan et al., explores the barriers clients face in adopting public cloud computing (PCC). Through a systematic literature review (SLR), the study identifies 29 challenges, with 19 deemed critical, including lack of security, privacy concerns, and data loss risks. Analyses reveal variations in challenges across time, continents, and publication venues. The findings aim to guide organizations and vendors in addressing these obstacles to enhance PCC adoption.**Citation (Chicago Style):**  
   Khan, Muhammad Janas, Fasee Ullah, Muhammad Imran, Jahangir Khan, Arshad Khan, Ahmed S. AlGhamdi, and Sultan S. Alshamrani. "Identifying Challenges for Clients in Adopting Sustainable Public Cloud Computing." *Sustainability* 14, no. 16 (2022): 9809. <https://doi.org/10.3390/su14169809>.
2. The paper "Energy-Efficient Resource Allocation and Migration in Private Cloud Data Centre" by Senthamil Selvan et al. proposes energy-efficient strategies for resource allocation and virtual machine (VM) migration in private cloud data centers. It introduces a Linear Numerical Programming (LNP) algorithm to optimize VM placement, reduce active servers, and minimize energy consumption. The study highlights the role of precise VM allocation and migration techniques in improving data center efficiency, reducing costs, and addressing challenges like power usage and resource management. Future work suggests integrating machine learning for enhanced optimization.

Citation (Chicago Style)

Selvan, Senthamil, Dhaya R., Daniel Krah, and Kalidoss Rajakani. "Energy-Efficient Resource Allocation and Migration in Private Cloud Data Centre." Journal of Cloud Computing 2022. <https://doi.org/10.1155/2022/3174716>.

1. <https://ieeexplore.ieee.org/abstract/document/9788254>
2. <https://ieeexplore.ieee.org/abstract/document/9467912>
3. <https://ieeexplore.ieee.org/abstract/document/9673425https://ieeexplore.ieee.org/abstract/document/9673425>