*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 1, Ver. I (Jan.-Feb. 2017), PP 95-99*
*www.iosrjournals.org*

# Computational approaches to humanities research: ethical and social implications of technology

Shovan Roy*
*Assistant Professor, Department of Computer Science, Midnapore College (Autonomous),*
*E-mail : sho.cmsa.08@gmail.com*
**Corresponding author*

***Abstract***
*During the past few years, we have seen an increase in the popularity of computational methods in humanities research, which use computer science, data analysis, and artificial intelligence (AI) techniques to explore humanistic fields. With the use of these methods, academics now have new ways to examine and make sense of large amounts of data, spot trends, and learn more about many facets of human society, history, language, and other topics. But, like with any technological development, there are moral and social ramifications to take into account. I will discuss here different aspects of humanities research under the umbrella of the computer science discipline.*
***Keywords:*** *Computer Science, Humanities research, Technology, Data analysis, Artificial Intelligence*

## I. Introduction

### 1.1 Motivation

Interdisciplinary cooperation, critical thought, and continual discussion among academics, politicians, and society at large are necessary to address these ethical and societal ramifications. Researchers can use computational methods in the humanities while respecting moral principles and advancing responsible and inclusive scholarship by considering these issues.

**Privacy and data protection:** Computational methods frequently call for access to massive datasets that contain sensitive data. Privacy and data protection must be ensured to avoid improper usage, reduce the possibility of re-identification, and safeguard people's rights. Anonymizing data, creating secure data storage and sharing procedures, and diligently collecting informed consent are all requirements for researchers.

**Fairness and Bias:** Computational approaches rely significantly on models and algorithms that may induce biases. Prejudices in society may be maintained, or some groups may be marginalized as a result of biased training data or algorithmic conclusions. By tackling discrimination in data collecting and algorithm design, it is critical to be conscious of such biases and actively strive toward establishing inclusive and equitable computational methodologies.

**Interpretation and Representation:** Critical analysis and subjective interpretation are frequent components of humanities studies. Although computational methods can find patterns and connections, human judgment is still needed for the interpretation of the findings. To prevent oversimplification or reductionism, it's critical to strike a balance between quantitative analysis and qualitative understanding. Researchers should consider how computational techniques affect how cultural, historical, and social phenomena are represented.

**Intellectual property and copyright:** Copyrighted assets, such as digitized texts, photographs, or audiovisual information, may be used in computational procedures. In order to use data, researchers must secure the required permits and be conscious of intellectual property rules. Fair use guidelines and open access efforts can aid in enabling responsible and authorized access to digital resources.

**Algorithmic Accountability and Transparency:** Transparency and accountability in algorithms are issues brought up by using sophisticated algorithms and AI systems in humanities studies. Researchers should work to explain algorithmic decisions, document their choices, and use transparent computational methodologies. Reproducibility, examination, and the ability to rectify potential biases and errors are all made possible through transparency.

**Impact on Traditional Research Methods:** Computational methodologies may affect traditional humanities research. Computational methods may have benefits, but they also have drawbacks, and researchers should be aware of both to ensure that new ways add to existing research procedures rather than replace them. Interdisciplinary research can be successful when computational professionals and domain experts collaborate.

**Technological and Access Divides:** Digital literacy, technical expertise, and access to computing resources are all necessary for computational techniques. Disparities in research opportunities and results can be caused by unequal access to technology and resources. It's critical to take inclusion into account and make sure computational methods don't widen already-existing technical gaps.

**Sustainable Long-Term Preservation:** The long-term preservation and sustainability of research outputs are challenged by the rapid evolution of digital resources and computational techniques. Adopting best practices for data management, ensuring interoperability and data standards, and taking into account the long-term availability and usability of their computational tools and models are all important considerations for researchers.

Interdisciplinary cooperation, critical thought, and continual discussion among academics, politicians, and society at large are necessary to address these ethical and societal ramifications. Researchers can use computational methods in the humanities while respecting moral principles and advancing responsible and inclusive scholarship by taking into account these issues.

**1.2 Literature review**

Numerous computational techniques can be applied in the humanities while upholding moral standards and promoting ethical and diverse scholarship. Here are a few instances:

i) **Natural language processing (NLP) and text mining (TM):** Text mining and NLP are techniques for examining and extracting data from massive textual databases. Language patterns, sentiment analysis, topic modelling, and entity recognition can all be researched using NLP techniques. Researchers can respect ethical norms while gaining valuable insights from textual data by ensuring that data anonymization and privacy are protected appropriately.

Text mining study [1] is becoming one of the most popular areas for examining texts written in natural language. The current paper provides a thorough summary of text mining's history and current state of research. The use of Data Mining techniques to address Information Extraction from research publications has limitations, as noted in the literature. Their interplay enables the discovery of several intriguing retrieved articles with various text patterns. In this work, they gathered 300 peer-reviewed journal articles on mobile-learning from six scholarly databases, including Wiley, Cambridge, Science Direct, IEEE, SAGE, and Springer. We textually evaluated them using various text-mining approaches. The criteria used to choose the papers for the collection was that they all had to emphasize mobile-learning for the higher education. According to experimental findings, the Springer database serves as the primary repository for research publications in the area of mobile-learning for the medical and pharmacy sector.

Additionally, when similarities between themes could not be recognized, it was either because of their interrelationships or because of ambiguity in their meaning. Additionally, research revealed that the number of publications published from 2015 to 2016 increased dramatically. The study also presents further implications and prospectives for the future.

In most businesses today, unstructured models are used to store information. In semantic web applications, information retrieval and extraction are crucial tasks. The storage effectiveness and analysis of unstructured data will determine many of these requirements. According to a recent estimate by Merrill Lynch, unstructured data makes up more than 80% of all potentially relevant corporate information. The analyst has a wide range of new options because of the quantity and complexity of unstructured data. Both individually and collectively, they examine structured and unstructured data [2]. NLP and TM are two ways to extract knowledge from textual content in articles/documents. TM and NLP strategies will be demonstrated in this study. This work aims to compare and assess the similarities/dissimilarities and differences between TM and NLP to extract relevant information using the best techniques available.

The rise of biomedical industry based literature drives the demand for effective tools [3] to keep up with the growing volume of information. It is becoming increasingly important to apply text mining (TM) techniques to enable the automatic extraction of valuable biomedical data from unstructured/unorganized text. They have examined the uses of TM in psychological field and looked at both its benefits and drawbacks. A thorough evaluation of the literature was conducted using databases from Medline, CINAHL, EMBASE, Cochrane, and PsycINFO. Among the 1103 papers that were examined for inclusion in this study as TM applications in psychiatric research, 38 were chosen. TM and content analysis allowed us to pinpoint four key application areas:

1. Studying mental disorders from an observational perspective (psychopathology)
2. Perspectives of the patient (that is, the opinions and thoughts of the patient)
3. Medical diagnosis records (i.e., safety concerns, care standards, and treatment descriptions)
4. Identifying new scientific information in the medical literature (i.e., research in the medical literature)

A variety of sources of information were used, including qualitative studies, medical records, Internet postings, and biomedical literature. Their work indicates how TM can help with challenging psychiatric research projects. They have discussed this instrument's advantages, restrictions, and potential future uses.

ii) **Network Analysis:** Mapping linkages and interactions between people, groups, or concepts is the goal of network analysis. It can be used to research social networks, academic partnerships, citation trends, and cultural sway. When employing network data, researchers should consider privacy, permission, and data protection concerns to ensure that people's identities and interpersonal connections are effectively safeguarded.

The main concepts, nature, and purpose of social network analysis are introduced to the non-specialist reader in this book [4]. Social networks function on various scales, from the level of families to that of entire nations, and are crucial in deciding how issues are resolved, how businesses are operated, and how well people can accomplish their goals. The social network theory depicts these connections among different actors. Despite being relatively new, it has significantly impacted the social sciences. This book introduces the fundamental concepts in context using examples and pictures, assuming no prior understanding of quantitative sociology. John Scott provides recommendations for further reading and online sources using a systematic approach to comprehending the work in this field, allowing readers to advance their knowledge and expertise to become practitioners of this research methodology.

The physical and social sciences have seen a surge in interest in network research during the past ten years. The idea of networks has been a gold mine for social scientists, providing answers for social phenomena in fields ranging from psychology to economics. Here, researchers have discussed [5] the types of phenomena that social scientists have attempted to explain using social network analysis and briefly summarise the fundamental presuppositions, objectives, and explanatory frameworks common in the discipline. They are working with academics from different areas of the physical and social sciences to better understand the causes and impacts of network phenomena.

iii) **Image analysis and computer vision:** Computer vision techniques allow for analyzing and interpreting visual materials like pictures, paintings, or old documents. Researchers can use algorithms for image analysis for tasks like image categorization, object recognition, or style analysis. When using visual resources, it is crucial to guarantee proper sourcing and copyright licenses.

Visual Question Answering (VQA), a task that is free-form and open-ended, has been suggested by researchers [6]. The aim is to offer an accurate natural language response to a natural language inquiry concerning a picture that is given. As in real-world situations like assisting blind people, both the questions and the answers are open-ended. Visual inquiries focus on specific aspects of an image, such as foreground details and underlying context. Because of this, a system that excels at VQA often requires a more in-depth comprehension of the image/picture and sophisticated reasoning than a system that generates generic picture descriptions.

Additionally, because many/multi open-ended responses/answers either comprise a few words or a limited number of options that can be offered in a multiple-choice style, VQA is suited to computer review. They offer a dataset of 0.25 million photos, 0.76 million questions, and 10 million responses (www.visualqa.org) and talk about the data it contains. Various VQA base-lines are offered to evaluate against human performance.

iv) **Geospatial Analysis:** Examining geographical and geographic patterns and interactions is the focus of geospatial analysis. Researchers can use geographic information systems (GIS) and satellite data to analyze historical landscapes, cultural geography, or urban growth. When working with location data, it's important to consider privacy concerns and ensure the information is suitably anonymized and aggregated.

Modern information & communication technologies (ICT) [7] are altering travel and activity designs/patterns, which may have substantial implications for how we live our daily lives and how we organize space. To investigate the complex spatiotemporal linkages between activities and interactions occurring in real/physical and virtual/social-network domains, time geography, which looks at human activities/functions under different restrictions/constraints in a space-time context, which provides a valuable framework. However, the characteristics of virtual functions and interactions carried out via ICT can not be adequately described and examined within the traditional time-geographic paradigm. In order to sufficiently capture and analyze all actions and interactions in a hybrid physical-virtual space, this article extends traditional time-geographic ideas.

v) **Data visualization:** By presenting complex data in a visual style, data visualization techniques make it easier to understand and analyze the information. Interactive visualizations can help researchers engage larger audiences and effectively share their findings. It is crucial to ensure that visualizations are created with accessibility in mind, without misrepresenting or interpreting data, and with a focus on inclusivity.

R is quickly gaining ground as the preferred environment for data analysis and visualization in both academia and business [8]. Lattice significantly increases R's functionality by bringing the tried-and-true Trellis graphics design, which William S. Cleveland and colleagues at Bell Labs first created for S. Lattice is a robust and beautiful high-level data visualization system that is enough for the majority of daily graphics needs while being easily extensible to meet the needs of cutting edge research. This book, written by the creator of the lattice system, goes into great detail on it, starting with the fundamentals and progressively going into particular low-level features as needed. The book can be read without any prior knowledge of latticework.

vi) **AI and Machine Learning (ML) for Classification and Prediction (CP) :** ML algorithms can help with categorization, classification, and outcome prediction based on data patterns. In order to avoid algorithmic biases, researchers should work toward fairness and transparency when developing models. Maintaining ethical standards can be achieved by ensuring varied and representative training datasets and routine model audits.

The SMAC (Social, Mobile, Analytic, Cloud) technological movement is currently paving the path/way for a advance future, in which extensive data, networked procedures, and intelligent machines will all operate together. The massive volume of data produced/created by virtual/internet environment is pushing the use of ML techniques and solutions. Computers are now able to mimic the human behaviour for ML. As the system learns from each interaction and activity, it can use it as experience for the next time.

vii) **Digital archives and preservation:** Computational techniques can help develop and upkeep digital archives, assuring the long-term preservation and accessibility of materials from cultural heritage. When digitizing and sharing archive documents, researchers should take into account concerns with intellectual property, copyright, and open access while preserving the rights of creators and communities.

The management of digital data, including its preservation for use in the future, is known as digital curation [9]. Information professionals actively participate in this process. Data curation and digital preservation have been practiced for decades, but recent developments have brought together a number of concepts, groups, and individuals to focus more on digital preservation. The National Science Foundation, the American Council of Learned Societies, and Dr. Liz Lyon, UKOLN, have all published reports highlighting the elements of digital curation that must be in place in order to maintain, preserve, and make digital objects accessible for use in the future. These findings, as well as the increased emphasis on research and the introduction of new insightful educational initiatives, have helped digital curation to take off and given digital curators a new path/way into the information technology professions. As digital asset management, data curation, electronic records management, and digital preservation become more and more popular, digital curation becomes a catch-all term.

viii) **Computational Ethnography:** Computational approaches can assist ethnographic research by examining extensive social media data, online groups, or digital cultural behaviours. Informed consent, privacy, and the ethical handling of personal data are all ethical issues. Researchers should be aware of potential biases and cultural sensitivity when evaluating digital behaviour and social dynamics.

Millions of social media users worldwide publish an increasing volume of content every day [10]. The majority of this stuff is accessible, searchable, and public. The number of papers offering various strategies and algorithms/methodologies to use the content as data science researchers in various fields is also increasing at the same time. The existing studies using Twitter, Facebook, Instagram, and YouTube as data/information sources are framed in this article's present literature assessment in the context of the classic social science methods of

ethnography, statistics, and computation. The objective is to provide/supply a summary of the advantages/merits and disadvantages of various strategies in light of the opportunities provided by various platforms.

MySpace and Facebook, among other social networking sites, act as "networked publics." Youth use networked publics, like unmediated publics like parks and malls, to assemble, interact with their peers, and understand and contribute to the culture around them. A study conducted by the researchers [11] examines how American adolescents connect with one another through networked publics, as well as characteristics of these mediated environments, such as persistence, searchability, and replication. In order to examine how young people perceive these structural characteristics and come up with creative solutions to make these systems function, ethnographic data are used. The activities of teenagers are used to analyze issues like privacy and impression control, and youth engagement in social networking sites is set within a historical context.

**2. Conclusion and Future Scope**

It is vital for researchers to follow ethical rules, gets the required approvals, takes into account any potential biases and limitations of algorithms, and engage in critical reflection throughout the study process when using any of these computational tools. Researchers can progress humanistic studies while respecting moral beliefs by using these approaches ethically and inclusively.

Researchers also who is highly interested in the computer science domain and the application of NLP, TM, databases and so on should know and enhance their knowledge about the fundamental technique/algorithm and tools of computer science to break the barrier between a particular language and a computer system, and be included into a cross-platform which progress the dimension of research in all aspects for the sustainable society.

## References

[1]. Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2017). A survey of text mining in social media: facebook and twitter perspectives. Adv. Sci. Technol. Eng. Syst. J, 2(1), 127-133.
[2]. Gharehchopogh, F. S., & Khalifelu, Z. A. (2011, October). Analysis and evaluation of unstructured data: text mining versus natural language processing. In 2011 5th International Conference on Application of Information and Communication Technologies (AICT) (pp. 1-4). IEEE.
[3]. Abbe, A., Grouin, C., Zweigenbaum, P., & Falissard, B. (2016). Text mining applications in psychiatry: a systematic literature review. International journal of methods in psychiatric research, 25(2), 86-100.
[4]. Scott, J. (2012). What is social network analysis? (p. 114). Bloomsbury Academic.
[5]. Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. science, 323(5916), 892-895.
[6]. Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C. L., & Parikh, D. (2015). Vqa: Visual question answering. In Proceedings of the IEEE international conference on computer vision (pp. 2425-2433).
[7]. Shaw, S. L., & Yu, H. (2009). A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. Journal of Transport Geography, 17(2), 141-149.
[8]. Sarkar, D. (2008). Multivariate data visualization with R. Use R.
[9]. Yakel, E. (2007). Digital curation. OCLC Systems & Services: International digital library perspectives, 23(4), 335-340.
[10]. Giglietto, F., Rossi, L., & Bennato, D. (2012). The open laboratory: Limits and possibilities of using Facebook, Twitter, and YouTube as a research data source. Journal of technology in human services, 30(3-4), 145-159.
[11]. Boyd, D. (2008). Why youth (heart) social network sites: The role of networked publics in teenage social life. YOUTH, IDENTITY, AND DIGITAL MEDIA, David Buckingham, ed., The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning, The MIT Press, Cambridge, MA, 2007-16.