Human-Computer Interaction: A Literature Analysis from to Using Automated Text Mining

Janhvi A. Chauhan[1], Dhaval M. Modi[2],
[1,2]*Lecturer, Government Polytechnic, Ahmadabad.*

***Abstract*—This study aims to explore the research themes and trends in the field of Human-Computer Interaction (HCI) over the past two decades, specifically from to. To achieve this, an automated content analysis utilizing text mining and probabilistic topic modeling was conducted on 33,830 peer-reviewed journal articles published between and, indexed in the SCOPUS database. The analysis revealed 53 distinct topics that define the themes and trends within the HCI field. These topics were further categorized, leading to the development of a systematic taxonomy that maps the evolution of HCI research and practices over the past 20 years. This taxonomy comprises six primary categories: "machinery systems," "HCI body of knowledge," "feature identification," "brain-computer interfaces," "interaction," and "medical." This study is expected to offer valuable insights and contribute to the advancement of HCI research, practices, and investments in the future.**

***Index Terms*—human-computer interaction, research themes and trends, literature analysis, text mining, topic modeling**

I. INTRODUCTION

Human-Computer Interaction (HCI) is an interdisciplinary area of research and practice focused on the design, application, and evaluation of interactive systems and technologies. A specific definition of HCI is provided by Hewett et al. (1992, p. 5), which states: "Human-computer interaction is a discipline concerned with the design, evaluation, and implementation of interactive computing systems for human use, along with the study of major phenomena surrounding them." As HCI explores the interaction between humans and computers, it integrates knowledge from various fields, including human behavior, psychology, sociology, cognition, anthropology, and education, as well as disciplines related to computer technologies, software engineering, ergonomics, industrial design, and graphic design.

In addition to its interdisciplinary nature, HCI is a dynamic and continuously evolving field, heavily influenced by technological advancements. This has led to a growing interest in HCI research. The first study in this domain was conducted by Shackel in 1959. Following this, the field began to develop further with the launch of the *International Journal of Man-Machine Studies in 1969. During the 1970s, the concept of "user-friendliness" was introduced, and in 1976, NATO sponsored a workshop on man-computer interaction. In 1982, the ACM (Association for Computing Machinery) established SIGCHI (Special Interest Group on Computer-Human Interaction) and began organizing the CHI (International Conference on Human-Computer Interaction). In the late 1980s, four significant publications were released in the field of HCI [5]. During this time, the first industrial laboratories dedicated to HCI began to emerge. With the advancement of responsive web technologies in the 1990s, the importance of HCI grew, shifting the focus toward user-centric designs rather than technical or designer-centric approaches.

Advancements in information and communication technologies have had a direct impact on HCI studies. As a result, HCI has been emerging as a distinct discipline since the late 1970s [5]. The field continues to attract growing interest from both researchers and practitioners across various disciplines. With the rapid expansion of user-responsive information technologies, research and practical applications in HCI have steadily increased each year. For instance, while the total number of journal articles published between 2000 and 2010 was 10,858, this figure rose to 22,972 in the following decade, from 2009 to 2018.

From this perspective, investigating and understanding emerging themes and trends in HCI research is strategically important for both the present and future of the field. Given the vast number of studies within HCI, analyzing all of them to gain a comprehensive overview is challenging. There are relatively few literature reviews that explore the developmental stages and paradigm shifts within the domain. Such reviews are essential for identifying research gaps, understanding trends, and guiding future studies.

In this context, Li and Zhang (2005) conducted a review of 337 HCI studies published between 1990 and 2002 across seven HCI-specific journals [6]. Expanding on this work, Zhang et al. (2009) analyzed 758 articles published between 1990 and 2008, classifying HCI research into IT development, IT use and impact, and general research topics. They also categorized study methodologies as empirical or non-empirical and identified research fields spanning information, computing, and communication sciences, as well as information systems [7].

Additionally, Agrawal et al. (2010) examined cultural issues, enterprise system involvement, and levels of analysis within HCI research, highlighting a gap in dyadic communication studies [8]. From a similar perspective, Kjeldskov and Paay (2012) conducted a review of 144 studies focusing on mobile HCI research methods [9]. Their findings indicated that mobile HCI research was primarily conducted empirically and later evolved to incorporate methodologies that integrated approaches from multiple disciplines [9]. Another meta-review by Coursaris and Bontis (2012) examined research productivity in the HCI field by analyzing publications from three key domain-specific journals [10].

However, since these and similar review studies were conducted on a limited number of publications, within restricted time frames, or using specific publication sources, a significant gap remains in the literature for capturing the broader landscape of HCI research. To address this, this paper employs an automated text-mining-based content analysis using Latent Dirichlet Allocation (LDA) to investigate research themes and trends in HCI. This topic modeling analysis was conducted on 33,830 journal articles published between 2000 and 2020. As a result, 53 trending topics were identified, offering a detailed insight into the evolving research directions in HCI.

The findings of this study are expected to provide valuable insights for the HCI industry, academic researchers and institutions, as well as the broader HCI community.

II. RESEARCH METHODOLOGY

A. Data Collection

The data collection process began with retrieving HCI-related articles from Scopus, a comprehensive bibliographic database that organizes scholarly information in various formats. To extract relevant articles, the following search query was applied in Scopus:

TITLE-ABS-KEY (("human computer" OR "computer human" OR "machine human" OR "human machine" OR "human robot" OR "robot human" OR "brain computer" OR "computer brain" OR "man machine" OR "machine man") AND (interact OR interface OR system*)) AND (PUBYEAR >) AND (PUBYEAR <) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "re"))
AND (LIMIT-TO (LANGUAGE, "English")).

This query was designed to retrieve only research and review articles published in English within a 20-year period, from to. As a result, a total of 33,830 journal articles containing the specified keywords in their title, abstract, or author keywords were identified. These articles formed the experimental dataset, which included metadata such as the title, abstract, and author keywords for each publication.

B. Data Preprocessing

Text preprocessing is a crucial step in text mining, aimed at cleaning and structuring complex textual data [11]. It significantly enhances text mining processes by improving analysis sensitivity, diagnostic accuracy, dimensionality reduction, and feature selection.

In this study, the preprocessing workflow involved several key steps:

- Tokenization

- Removal of numbers, punctuation, and links
- Conversion of text to lowercase
- Elimination of stop words
- Application of stemming using the Porter Stemmer algorithm [12, 13]

These preprocessing steps facilitated the creation of a lower-dimensional document-term matrix (DTM), which effectively represented the dataset [13, 14].

C. Topic Modeling Analysis with LDA

This study utilized Latent Dirichlet Allocation (LDA) [15], a generative probabilistic topic modeling technique, to uncover latent semantic structures within the HCI articles in the dataset. In LDA, each document is represented as a probabilistic mixture of topics, while each topic is defined by a distribution of words [15, 16]. As an unsupervised machine learning method, LDA enables the rapid analysis of large textual datasets, making it a widely adopted approach in text mining research [16-20].

For implementing LDA on the experimental dataset, the MALLET package [21], a Java-based topic modeling tool, was employed. The model was executed with 2,600 Gibbs sampling iterations to optimize topic discovery [22]. LDA relies on two Dirichlet hyperparameters, α and β, to fine-tune the model [15, 23]. Additionally, the number of topics (T) is a key parameter that controls the granularity of the topic distribution [16, 23]. In this study, LDA was tested with different T values ranging from 20 to 75, with the most meaningful topic distribution achieved at $T = 53$. Once the topics were identified, they were manually labeled based on the most frequently occurring keywords within each topic [17, 18].

III. RESULTS

This study analyzed the HCI research literature to identify underlying themes and trends from 2000 to 2020. By applying Latent Dirichlet Allocation (LDA) to an experimental dataset of 33,830 journal articles, 53 key topics representing HCI research trends were identified. Table I presents these topics along with their associated keywords and respective frequencies, listed in descending order.

The findings reveal not only the evolving research directions in HCI but also the field’s application areas and emerging systems. Among the identified topics, "user interfaces" emerged as the most prominent research focus, reflecting its central role in the HCI discipline. "Requirements" ranked second, followed by "research" in third place. Other frequently studied topics include "machine-human systems," "software development," "cognition," "nurse training," "user experience," "feature classification," and "production control."

To further structure and present these findings in a more interpretable manner, a research taxonomy of HCI was developed. This taxonomy categorizes the 53 topics into six overarching research and application domains. The six main categories forming this taxonomy are:Machinery Systems, HCI Body of Knowledge, Feature Identification, Brain-Computer Interfaces, Interaction, Medical ApplicationsTable II presents the distribution of these topics across the six categories along with their corresponding percentages, providing a clearer perspective on HCI research trends.

The first research category, Machinery Systems (25.26%), includes 14 topics, covering various aspects of machine-human interaction and control. These topics are: machine-human systems, production control, computer control systems, human-robot interaction, mobile, force control, intelligent decision support, display device, human safety, robot control, sensors, flight control systems, driving control systems, and energy flow.

The second category, HCI Body of Knowledge (21.04%), comprises seven key topics that form the foundational aspects of HCI research: user interfaces, requirements, research, software development, user experience, usability evaluation, and user.

The third category, Feature Identification (19.45%), consists of 11 topics focusing on classification, recognition, and data processing techniques. These include: feature classification, response time, image detection, data-knowledge, gesture, information retrieval, optimization algorithms, natural language processing, neural networks, speech recognition, and emotion.

The fourth category, Brain-Computer Interfaces (12.36%), encompasses six topics related to cognitive and neural interactions: cognition, cognitive tasks, EEG signals, brain-computer interface, brain-neural interface, and brain-motor movement.

The fifth category, Interaction (11.60%), includes seven topics centered on digital and virtual engagement: digital interaction, simulation models, education, social media, virtual reality, visual feedback, and video games.
Lastly, the sixth category, Medical Applications (10.82%), consists of eight topics related to healthcare and medical advancements: nurse training, healthcare, disability rehabilitation, musculoskeletal disorders, disease intervention, genetics, surgery, and animal tissue research.

TABLE I. RESEARCH TOPICS DISCOVERED BY LDA.

ID	TopicName	DescriptiveKeywords	Rate (%)
1	Userinterfaces	comput*;user*;interfac*;interact*;human*;system*;graphic*;design*;applic*;base*	3.84
2	Requirements	approach*;requir*;applic*;cost*;limit*;provid*;techniqu*;time*;effect*;complex*	3.72
3	Research	research*;technologi*;review*;develop*;human*;comput*;futur*;discuss*;field*;issu*	3.68
4	Machine-humansystems	system*;machin*;human*;oper*;inform*;engin*;analysi*;scienc*;develop*;cybernet*	3.56
5	Softwaredevelopment	design*;develop*;softwar*;tool*;engin*;process*;comput*;product*;interact*;project*	3.18
6	Cognition	cognit*;human*;interact*;action*;theori*;framework*;activ*;comput*;process*;context*	3.00
7	Nursetraining	comput*;nurs*;train*;educ*;human*;attitud*;inform*;health*;manag*;medic*	2.81
8	Userexperience	experi*;person*;self*;behavior*;effect*;particip*;perceiv*;research*;result*;user*	2.53
9	Featureclassification	featur*;classif*;pattern*;classifi*;analysi*;extract*;recognit*;method*;vector*;machin*	2.51
10	Productioncontrol	control*;system*;oper*;machin*;product*;manufactur*;industri*;process*;plant*;interfac*	2.40
11	Cognitivetasks	task*;perform*;human*;mental*;cognit*;autom*;workload*;effect*;experi*;control*	2.37
12	Digitalinteraction	interact*;digit*;design*;technologi*;comput*;human*;hci*;cultur*;creativ*;experi*	2.36
13	Usabilityevaluation	evalu*;usabl*;test*;method*;measur*;qualiti*;assess*;user*;result*;valid*	2.30
14	Responsetime	time*;rate*;respons*;measur*;result*;frequenc*;real*;chang*;condit*;subject*	2.28
15	Imagedetection	imag*;detect*;comput*;method*;object*;color*;human*;vision*;base*;segment*	2.24
16	Computercontrolsystems	system*;comput*;control*;data*;softwar*;interfac*;network*;commun*;time*;power*	2.21
17	Human-robotinteraction	robot*;human*;interact*;humanoid*;social*;behavior*;assist*;hri*;anthropomorph*;physic*	2.20
18	Simulationmodels	model*;simul*;predict*;comput*;human*;dynam*;base*;mathemat*;analysi*;method*	2.18
19	Mobile	mobil*;servic*;comput*;user*;devic*;network*;smart*;secur*;applic*;interact*	2.16
20	Forcecontrol	control*;forc*;human*;mechan*;system*;dynam*;motion*;actuat*;joint*;	2.10

		design*	
21	Intelligentdecisionsupport	intellig*;decis*;agent*;system*;support*;human*;collabor*;artifici*;comput*;cooper*	1.95
22	Data-knowledge	data*;knowledg*;base*;expert*;semant*;inform*;analysi*;process*;map*;structur*	1.95
23	Gesture	gestur*;track*;hand*;ey*;human*;motion*;recognit*;movement*;interact*;gaze*	1.92
24	Informationretrieval	inform*;web*;search*;retriev*;user*;internet*;access*;content*;onlin*;websit*	1.90
25	EEGsignals	signal*;brain*;eeg*;interfac*;comput*;bci*;process*;electroencephalographi*;channel*;nois*	1.89
26	Education	learn*;student*;educ*;comput*;teach*;instruct*;base*;cours*;interact*;skill*	1.83
27	User	ag*;adult*;femal*;male*;comput*;internet*;human*;children*;adolesc*;middl*	1.79
28	Brain-computerinterface	brain*;bci*;interfac*;comput*;potenti*;visual*;evok*;base*;event*;relat*	1.76
29	Brainneural interface	brain*;interfac*;neural*;comput*;cortex*;function*;stimul*;neuron*;activ*;decod*	1.71
30	Health-care	health*;medic*;care*;patient*;inform*;record*;clinic*;system*;electron*;physician*	1.68
31	Socialmedia	social*;commun*;onlin*;media*;network*;behavior*;trust*;interact*;human*;mediat*	1.67
32	Brain-motor-movement	motor*;movement*;brain*;imageri*;interfac*;bci*;comput*;control*;hand*;perform*	1.63
33	Displaydevice	displai*;devic*;screen*;touch*;comput*;input*;mous*;keyboard*;interact*;design*	1.56
34	Humansafety	error*;human*;safeti*;ergonom*;risk*;factor*;accid*;analysi*;occup*;worker*	1.55
35	Optimizationalgorithms	optim*;algorithm*;method*;comput*;function*;schedul*;solut*;decis*;interact*;select*	1.54
36	Naturallanguageprocessing	languag*;program*;comput*;natur*;text*;word*;commun*;process*;code*;charact*	1.49
37	Robotcontrol	robot*;control*;manipul*;oper*;navig*;plan*;remot*;task*;environ*;assembli*	1.47
38	Sensors	sensor*;electron*;devic*;wearabl*;human*;sens*;skin*;pressur*;materi*;flexibl*	1.41
39	Virtual-reality	virtual*;realiti*;environ*;interact*;simul*;augment*;comput*;real*;world*;human*	1.38
40	Disabilityrehabilitation	rehabilit*;assist*;disabl*;patient*;stroke*;wheelchair*;devic*;control*;function*;motor*	1.31
41	Neuralnetworks	network*;learn*;neural*;adapt*;fuzzi*;algorithm*;artifici*;train*;system*;base*	1.30
42	Speechrecognition	speech*;recognit*;dialogu*;voic*;multimod*;commun*;modal*;system*;interact*;convers*	1.27
43	Visualfeedback	visual*;feedback*;percept*;auditori*;music*;attent*;haptic*;sensori*;sound*;human*	1.25
44	Musculoskeletal	muscl*;hand*;human*;postur*;electromyographi*;bodi*;emg*;activ*;musculoskelet*;upper*	1.19

45	Diseaseintervention	diseas*;disord*;patient*;intervent*;clinic*;trial*;comput*;therapi*;control*;treatment*	1.05
46	Emotion	emot*;express*;facial*;recognit*;affect*;human*;interact*;comput*;physiolog*;featur*	1.04
47	Flightcontrolsystems	pilot*;control*;flight*;aircraft*;space*;system*;oper*;air*;machin*;autom*	1.01
48	Video-game	game*;video*;plai*;player*;comput*;interact*;experi*;graphic*;seriou*;engag*	0.93
49	Drivingcontrolsystems	drive*;driver*;vehicl*;system*;car*;control*;simul*;automobil*;steer*;traffic*	0.87
50	Genetics	sequenc*;protein*;databas*;gene*;analysi*;comput*;genet*;data*;genom*;softwar*	0.85
51	Energy flow	energi*;flow*;environment*;water*;power*;effici*;temperatur*;consumpt*;grid*;electr*	0.81
52	Surgery	surgeri*;imag*;surgic*;assist*;medic*;comput*;procedur*;cancer*;surgeon*;diagnost*	0.74
53	Animaltissue	anim*;cell*;tissu*;blood*;drug*;probe*;experiment*;rat*;laser*;effect*;	0.65

TABLE II. RESEACH TAXONOMY FOR HCI

Category	TopicName	Rate(%)	Total(%)
Machinerysystems	Machine-humansystems	3.56%	25.26%
	Productioncontrol	2.40%	
	Computercontrolsystems	2.21%	
	Human-robotinteraction	2.20%	
	Mobile	2.16%	
	Forcecontrol	2.10%	
	Intelligentdecisionsupport	1.95%	
	Displaydevice	1.56%	
	Humansafety	1.55%	
	Robotcontrol	1.47%	
	Sensors	1.41%	
	Flightcontrolsystems	1.01%	
	Drivingcontrolsystems	0.87%	
	Energyflow	0.81%	
HCIbodyofknowledge	Userinterfaces	3.84%	21.04%
	Requirements	3.72%	
	Research	3.68%	
	Softwaredevelopment	3.18%	
	Userexperience	2.53%	
	Usabilityevaluation	2.30%	
	User	1.79%	
	Featureclassification	2.51%	
	Responsetime	2.28%	
	Imagedetection	2.24%	
	Data-knowledge	1.95%	
	Gesture	1.92%	
	Informationretrieval	1.90%	

Featureidentification	Optimizationalgorithms	1.54%	19.45%
	Naturallanguageprocessing	1.49%	
	Neuralnetworks	1.30%	
	Speechrecognition	1.27%	
	Emotion	1.04%	
Brain-computerinterfaces	Cognition	3.00%	12.36%
	Cognitive tasks	2.37%	
	EEGsignals	1.89%	
	Brain-computerinterface	1.76%	
	Brainneuralinterface	1.71%	
	Brain-motor-movement	1.63%	
Interaction	Digitalinteraction	2.36%	11.60%
	Simulationmodels	2.18%	
	Education	1.83%	
	Socialmedia	1.67%	
	Virtual-reality	1.38%	
	Visualfeedback	1.25%	
	Video-game	0.93%	
Medical	Nursetraining	2.81%	10.29%
	Health-care	1.68%	
	Disabilityrehabilitation	1.31%	
	Musculoskeletal	1.19%	
	Diseaseintervention	1.05%	
	Genetics	0.85%	
	Surgery	0.74%	
	Animaltissue	0.65%	

IV. CONCLUSION

This study employed an automated text mining-based content analysis to examine HCI literature using Latent Dirichlet Allocation (LDA), a probabilistic topic modeling approach. The analysis covered 33,830 peer-reviewed journal articles published between and, uncovering 53 key research topics that define the HCI research landscape. Among these topics, "user interfaces" emerged as the most extensively studied, followed by "requirements," "research," "machine-human systems," and "software development."

To further structure these findings, the identified topics were systematically categorized into a research taxonomy that reflects the evolution of HCI over the past two decades. This taxonomy consists of six core categories: Machinery Systems, HCI Body of Knowledge, Feature Identification, Brain-Computer Interfaces, Interaction, Medical Applications

The insights derived from this study provide valuable implications for the HCI research community, shedding light on emerging trends and potential research gaps. As HCI remains a rapidly evolving field, these findings can help guide future studies and explorations.

Furthermore, the proposed research methodology offers a scalable approach for automated literature analysis, not only in HCI but also in other research domains. Additionally, this methodology can be applied to various online community platforms that generate vast amounts of textual data, such as forums, social networks, and blogs.

REFERENCES

[1] G. J. Kim and F. Group, “Human – Computer Interaction Fundamentals and Practice,” Hum. Comput. Interact. Fundam. Pract., pp. 1–12, 2015.

[2] T. T. Hewett et al., ACM SIGCHI curricula for

human-computer interaction. ACM, 1992.

[3] J. Lazar, J. H. Feng, and H. Hochheiser, Research methods in human computer interaction. Morgan Kaufmann, 2017.

[4] B. Shackel, "Ergonomics for a computer," Design, vol. 120, pp. 36–39, 1959.

[5] K. Çağıltay, İnsanbilgisayaretkileşimivekullanılabilirlikmühendisliği: Teoridenpratiğe. Seçkin Yayınevi, 2018.

[6] [6] N. L. Li and P. Zhang, "The intellectual development of human computer interaction research: A critical assessment of the MIS literature (1990-2002)," J. Assoc. Inf. Syst., vol. 6, no. 11, p. 9, 2005.

[7] P. Zhang, N. Li, M. Scialdone, and J. Carey, "The Intellectual Advancement of Human-Computer Interaction Research: A Critical Assessment of the MIS Literature (1990-2008)," AIS Trans. Human Computer Interact., vol. 1, no. 3, pp. 55–107, 2009.

[8] A. Agrawal, M. J. Boese, and S. Sarker, "A Review of the HCI Literature in IS: The Missing Links of Computer-mediated Communication, Culture, and Interaction.," in AMCIS, 2010, p. 523.

[9] J. Kjeldskov and J. Paay, "A longitudinal review of Mobile HCI research methods," in Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services, 2012, pp. 69–78.

[10] C. K. Coursaris and N. Bontis, "A Meta Review of HCI Literature: Citation Impact and Research Productivity Rankings," Sighci 2012, vol. 9, pp. 1–5, 2012.

[11] A. N. Srivastava and M. Sahami, Text mining: Classification, clustering, and applications. CRC Press, 2009.

[12] M. Porter, "Snowball: A language for stemming algorithms," Snowball. 2001.

[13] C. C. Aggarwal and C. X. Zhai, Mining text data, vol. 9781461432. 2013.

[14] F. Gurcan, "Major Research Topics in Big Data: A Literature Analysis from 2013 to 2017 Using Probabilistic Topic Models," in 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018, pp. 1–4.

[15] D. M. Blei, B. B. Edu, A. Y. Ng, A. S. Edu, M. I. Jordan, and J. B. Edu, "Latent Dirichlet Allocation," J. Mach. Learn. Res., vol. 3, pp. 993 1022, 2003.

[16] D. M. Blei, "Probabilistic topic models," Commun. ACM, vol. 55, no. 4, pp. 77–84, 2012.

[17] T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proc. Natl. Acad. Sci., vol. 101, no. Supplement 1, pp. 5228–5235, 2004.

[18] F. Gurcan and N. E. Cagiltay, "Big Data Software Engineering: Analysis of Knowledge Domains and Skill Sets Using LDA-Based Topic Modeling," IEEE Access, vol. 7, pp. 82541–82552, 2019.

[19] F. Gurcan and C. Kose, "Analysis of Software Engineering Industry Needs and Trends: Implications for Education," Int. J. Eng. Educ., vol. 33, no. 4, pp. 1361–1368, 2017.

[20] F. Gurcan, "Extraction of Core Competencies for Big Data: Implications for Competency-Based Engineering Education," Int. J. Eng. Educ., vol. 35, no. 4, pp. 1110–1115, 2019.

[21] A. K. McCallum, "MALLET: A Machine Learning for Language Toolkit," Http://Mallet.Cs.Umass.Edu. 2002.

[22] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[23] H. M. Wallach, "Topic Modeling: Beyond Bag-of-Words," ICML, no. 1, pp. 977–984, 2006.