Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Holistic Approach to Studying Interdisciplinary Collaboration: Computer Science

Shrabani Ghosh

sghosh15@charlotte.edu

University of North Carolina at Charlotte

Research Article

Keywords:

Posted Date: February 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8694985/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Holistic Approach to Studying Interdisciplinary Collaboration: Computer Science

Shrabani Ghosh
Computer Science, University of North Carolina Charlotte, Charlotte, North Carolina, USA

Abstract

Research collaboration networks play a crucial role in shaping the evolution of scientific fields. In computer science, interdisciplinary research has created a significant impact. It has contributed to merging knowledge from one discipline to another, generating new innovations and enriching all involved. For example, Hadoop was developed to enable large-scale distributed data processing, inspired by Google's MapReduce and GFS. Similarly, modern VR headsets emerged from advancements in graphics, system design, and motion tracking technologies. In the past few years, AI-driven research has influenced various domains and brought revolutionary changes to computer science. Understanding the growth of these domains requires a different aspect of analysis rather than focusing on a single aspect.

To examine interdisciplinary collaboration, we investigate three key aspects: (i) Topic Modeling – identifying trending subdisciplinary topics and their evolution; (ii) Cross-Disciplinary Authors – detecting bridging researchers who facilitate knowledge exchange; and (iii) Industry Collaboration – analyzing institutional partnerships and their role in shaping research trajectories.

This study compares four major computer science disciplines—Human-Computer Interaction, Computer Vision, Software Engineering, and Data Mining—using DBLP and ORCID data. Through a temporal analysis of topics, we identify emerging and declining research trends. Additionally, we highlight interdisciplinary knowledge flows, the role of bridging authors, and the evolving involvement of academia, industry, and research centers. Our holistic analysis provides a comprehensive view of computing research communities, their structural evolution, and the factors driving their collective growth.

Author summary

To understand collaboration among subdisciplines within a research domain, we examined three key aspects that contribute to the overall growth of a field. We focused on subdisciplines in computer science to explore cross-disciplinary collaboration. Our findings reveal that knowledge exchange among disciplines has fostered an interdisciplinary research culture, benefiting all involved. Bridging authors—who connect different research areas—have played a crucial role in facilitating this exchange, significantly enriching the field. Moreover, the rapid advancement of computer science has been instrumental in the AI boom, driven not only by researchers but also by active investments from industries and research institutions.

1 Introduction

Nowadays, numerous papers are published daily, and the volume of published papers has significantly increased over the last few decades [1]. Researchers tend to publish more in the early stages of their careers and have increased co-authorship in recent decades. However, according to Fanelli et al. [2], this does not enhance the overall productivity of researchers throughout the century. Publication and citation trends depend on the research area and the culture of a research community. Some research communities are more active in publishing papers at conferences, while others prefer to publish papers in long-term journal full-fledged publications. For example, Vrettas & Sanderson (2014) [3] show that Computer Science (CS) values conference publications more highly than any other academic field of study. From working on a research topic to develop work and publish papers, time differs across disciplines within CS. Some research areas require expensive resources that need funding, while others can be less costly to pursue. For example, robotics in CS involves hardware-related research that requires sufficient funding, whereas fields that are more software-based may be significantly less expensive in terms of resources. Comparing the number of publications and citations cannot effectively gauge the importance and activity of a research topic. Additionally, the application purpose of research influences its demand; for instance, medical research [4] and autonomous system development industries [5] invest in robotics compared to academic research.

It is interesting to see the different aspects of research happening in the CS domain and how various disciplines have developed over time. There are many areas within CS, such as High-Performance Computing, Medical Image Analysis, Software Engineering, Text Mining, Vision, Microprocessor & Hardware, Security & Privacy, and Artificial Intelligence, among others. From the theoretical concepts in CS, several application-based research disciplines have emerged, including Computer Vision, Data Mining, and Human-Computer Interaction. To compare the diverse set of disciplines in CS, we employed a holistic approach that looks at various aspect of research disciplines rather than focusing on a single aspect. This approach emphasizes topics and trends of disciplines, collaboration patterns and culture, and the involvement of industries and research centers in these disciplines. It provides a broad view of where disciplines are headed and how they are impacting the CS field as a whole. For topic analysis, we accommodated a topic modeling approach to examine the top research topics currently being pursued and how these topics have evolved over the past decade. This information is crucial for researchers and newcomers to the field to understand what is trending today. We have observed that, across most disciplines, AI-related research is growing and gaining more significance. The collaborative culture also varies among disciplines; some fields have more researchers, leading to increased publications and collaborations, while others have tighter-knit groups. The involvement of industry and academia also demonstrates different levels of interest in research. Industries invest in research that serves their business interests, while academia conducts varying amounts of research across disciplines.

1.1 Background & Related Work

Topic analysis is important to understand the ongoing research in a given domain. It reveals which topics are trending and which are losing relevance over time. Various studies have analyzed research topics in different ways [6, 7], and the purpose of topic analysis varies [8, 9]. Ji et al. (2018) [6] show how medical education has evolved over different phases of time. Ghosh et al. (2014) [9] analyzed topics related to railroad safety on Twitter social media. There many different ways to do the topic modeling [10] such as algebric models [11], probabilistic model [12], neural models [13, 14]. Recently,

Large Language Models are giving high performance of topic modeling such as BERT [15], where document is represented as embedding vector. BERT excels at capturing contextual word embeddings rather than traditional global embeddings that assign fixed vectors to words regardless of their usage, leading to more nuanced and accurate semantic understanding.

In research, authors are essential for advancing a field, and their global collaboration fosters true innovations. Analyzing collaboration networks reveals the collaborative culture among authors in a domain. Examples of such collaboration studies include [16–21]. These studies focused on identifying influential and leading authors and institutions in fields such as Medical Science, tourism, seaports, biomedical sciences, and patent data. Research has also been conducted on subdisciplines of Computer Science collaboration networks [22–26]. This paper [25] developed a link prediction-based collaboration recommendation application; [22] shows the collaboration level and number of authors in future publications in CS literature; [24] illustrates the structure and dynamics of CS research collaboration; and [23] highlights the properties of collaboration networks as well as the dependence on star collaborators. Pham et al.(2011) [26] analyzed computer science (combined DBLP and CiteSeerX databases) development by visualizing knowledge networks over time. An important conclusion of the papers is that conferences shape the social structures of computer science, making it increasingly interdisciplinary as sub-disciplines connect. This work examines citation and collaboration subgraphs of venues using various SNA metrics. Huang et al. (2008) [27] conducted a longitudinal analysis at both the network and community levels, comparing collaboration patterns across six topical communities and predicting collaboration between pairs of authors. Lara-Cabrera et al. (2013) [28] work has provided temporal perspective several properties of the computer intelligence in games network at the macroscopic, mesoscopic, and microscopic levels using DBLP data. However, none of the aforementioned papers have studied different disciplines in total within recent years, especially after the AI boom, because, in the past few years, computer science has undergone significant changes. This is a very important era for computer science, marked by the invention of Large Language Models that will change the next 20-30 years of research and the trajectory of the field.

To address this limitation, we conduct a study of computer science disciplines in recent years using DBLP data to understand topics and trends, collaboration across disciplines, and bridge nodes of different disciplines since they serve as brokers of information between various communities and institutions. In our study, we qualitatively examine the current topics and trends in those disciplines and observe that domains are merging concepts from one to another. In interdisciplinary collaboration, researchers integrate knowledge from two or more disciplines [29]. Researchers from different disciplines collaborate, combining knowledge drawn from their respective fields to generate new insights. We also study the involvement of industries in current research. For example, which companies are interested in a particular discipline or, in general, interested in research rather than being selective. These will also indicate the future of companies and provide a standpoint from which we can compare in future years.

In our previous work, [30], we presented comparative analysis on co-authorship networks for two research domains extracted from Google Scholar. The extensive experiments show publication trends, influential authors, and affiliated organizations for Data Mining and Software Engineering. The network analysis indicates that the network's features are distinct from one another and exhibit small communities among the influential authors of a particular domain. That work intrigued us to delve deeper and conduct an extensible analysis, building a framework where any domain's dataset can be integrated for comparative analysis. We extended our dataset in terms of disciplines and number of publication venues to establish the outcome of the comparison

is justifiable.

1.2 Research Question

The main purpose of this study is to present an expanded version of our previous work. We incorporate a few more research questions that provide an overall view of the growing interest in a holistic approach. In this work, we tried to answer the following research questions.

- What are the top most research topics emerging between 2012 to 2025 among all research disciplines in computer science? Are there emerging interdisciplinary research areas forming within domains, such as AI-driven research topics?
- Who are the key players/or bridging nodes in interdisciplinary research work? Which disciplines have a higher concentration of such key contributors?
- Which institutions rank highest in publishing papers within these disciplines, and what is their distribution of participation and collaboration?
- Which universities, companies, and research centers are most active in collaboration and publication across these four disciplines?

1.3 Structure of the paper

The structure of this paper is as follows: the next section (Sect. 2) introduces the data collection and data processing procedures in detail. We then introduce topic modeling (Sect. 3) and describe how the textual data was pre-processed for topic modeling. The results are presented in Sect. 3 and 3.1. Sect. 3 is devoted to showcasing the top 20 topics across all disciplines and their distributions and Sect. 3.1 shows topics shifts. Sect. 4 describes the creation of the collaboration network and provides statistical information about the network, while Sect. 4.1 highlights the bridging nodes across disciplines. In Sect. 5, we present our findings on the collaboration of institutions within disciplines. The paper concludes with the "Discussion" section, where we evaluate the potential limitations of our approach and summarize our conclusions from the results.

All code (Python 3.13.1) needed to replicate our analyses are available at Github The dataset links are cited properly.

2 Data Collection

2.1 DBLP

We are interested in analyzing computer science research fields to identify emerging topics and understand the network structure of scholarly collaborations. By examining publication trends, topics over years, authors collaboration, and research communities, we aim to discover patterns that highlight the evolution and cultures of the fields.

To achieve this, we collected raw DBLP data [31], a comprehensive computer science bibliography, which includes metadata such as authors names and ORCID IDs, title of papers, time of publication, publication venue, crossref, keyid, and DOI link. The full dataset has papers meta data available from year 1951 to 2025. DBLP records a total of 14,682 publication venues (booktitles) and provides information on 3,629,331 published papers. For our analysis, we are interested in specific computer science research domains; therefore, we selected Data Mining, Software Engineering, Computer Vision, and Human-Robot Interaction. These four domains were chosen to provide a broad comparative perspective across different areas of computer science. Data Mining has

grown significantly to address the challenges posed by big data across various domains. Software Engineering is a well-established, foundational and a more mature field, that continues to evolve. Computer Vision has become integral to innovation, with applications in industries such as medical diagnostics and automation. Human-Computer Interaction has contributed to advancements in human-robot interaction, with applications in medical surgery, industrial manufacturing, and beyond. Each domain has many types of publication venues, such as conferences, workshops, and journal papers published in DBLP. Some of these publishing venues make it difficult to distinguish research related to the domains. Therefore, we picked a handful of publication venues that are distinct and representative of their respective fields. These venues serve as key platforms for researchers to contribute their work from around the world, shaping trends, collaborations, and advances within their domains. For Data Mining (DM), we chose major conferences CIKM, KDD, SIGIR, WWW, WSDM. For Software Engineering (SW), we selected ASE, ESEC/SIGSOFT FSE, COMPSAC, ISSTA, ICSE, and UIST. For Computer Vision (CV), we focused on ICCV, ICPR, WACV, and CVPR. For Human-Robot Interaction (HRI), we considered HRI, HRI (Companion), ROBIO, IROS, and ICRA. In all these conferences, we excluded workshop papers (e.g., CVPR Workshop papers) as they typically undergo a different review process and may not be considered full-length, archival publications.

In DBLP, the names of authors are not always unique, and they can be ambiguous. For example, we found nine authors named "Bo Wang" with different ORCID IDs. Additionally, not all author names are mapped to an ORCID ID, making it difficult to solely rely on the author count based on ORCID IDs. However, some names have a numerical identifier appended in DBLP, such as "Yongxin Zhu 0001" and "Yongxin Zhu 0002," although this identification does not cover all the ambiguous names.

2.2 ORCID

From the DBLP dataset, we found that a percentage of authors have available ORCID IDs. The authors who published papers in the very early years and have not been very active recent years usually do not hold an ORCID profile. Thus, considering a longer period does not provide us with the information we need. That is why, for ORCID data collection, we focused on the years from 2022 to the current year, as, in these recent years, having an ORCID profile has become mandatory for many conference publications. ORCID IDs provide information such as the author's name, history of affiliated institutions, qualifications, education, and publication details. Although most ORCID ID profiles are not fully completed, they typically include only names and affiliations. We exported 2024 ORCID database [32]. Table 1 shows the percentage of available ORCID IDs of authors for each venue from 2022 to 2024, ranked from highest to lowest.

We kept the dataset separate from 2012 to 2025 and from 2022 to 2025 for different analytical purposes. Our interest in exploring the details of each discipline's hot and emerging research topics led us to retain the dataset for an extended period, such as a decade, because understanding the high-level interests and trends in a research area requires a lengthy timeframe. Another aspect of our analysis focuses on authors' affiliation information to create various types of collaborations across domains, considering the dataset from 2022 to 2025.

3 Topic Modeling

To obtain a topical overview of research, we apply topic modeling on articles titles to identify emerging research trends and thematic patterns within each discipline. Titles

Booktitle	Percentage ORCID
HRI (Companion) (Human Computer Interaction)	70.18%
UIST (Software Engineering)	69.71%
WSDM (Data Mining)	69.00%
WWW (Data Mining)	66.64%
ESEC/SIGSOFT FSE (Software Engineering)	64.19%
ISSTA (Software Engineering)	62.92%
CIKM (Data Mining)	62.50%
HRI (Human Computer Interaction)	59.12%
KDD (Data Mining)	58.87%
SIGIR (Data Mining)	57.61%
ASE (Software Engineering)	54.53%
ICPR (Computer Vision)	53.69%
ICSE (Software Engineering)	53.19%
ICCV (Computer Vision)	45.58%
WACV (Computer Vision)	41.10%
ROBIO (Human Computer Interaction)	32.15%
COMPSAC (Software Engineering)	32.14%
IROS (Human Computer Interaction)	30.21%
CVPR (Computer Vision)	28.62%
ICRA (Human Computer Interaction)	28.48%

Table 1. Percentage of ORCID IDs available for authors on DBLP within year 2022 to 2025

provide a concise summary of a paper's focus, making them an effective source for extracting key topics. By analyzing these topics, we can uncover shifts in research interests, detect evolving subfields, and gain insights into the broader landscape of computer science research. We used BERTopic [33] for topic modeling, which categorized a collection of text documents into clusters of titles that were semantically similar to each other, thereby discovering latent topics in the collection of titles. Before applying topic modeling, we preprocessed the titles, which includes converting to lower case, removing extra spaces, numbers, punctuation, and stop words, as well as stemming, which involves language-specific rules to remove or replace suffixes and prefixes. After preprocessing, we apply BERTopic, in that case we used "all-mpnet-base-v2" sentence model that extracts embeddings, a UMAP model that reduce dimensionality, an HDBSCAN model that clusters on reduced embeddings, a vectorization model (CountVectorizer), a CTFTIDF model (ClassTfidfTransformer), and a representation model (KeyBERTInspired). CountVectorize is used to tokenize topics. Each topic was represented with topmost important words relevant to that topic and each topic has atleast 40 papers.

We considered articles from 2012 to the most recent for the selected venues, which covers nearly a decade of research. Figure 1 shows the number of articles published for the selected conferences we mentioned earlier. The distribution illustrates the total count of articles for each research discipline per year. From figure, we can observe that the number of publications increases each year compared to the previous year, although some years may have slightly higher counts than others. In DM, we can see that 2021 has a higher number of publications, SW has consistently seen an increase in publications, CV had a higher number in 2023, and HRI had more publications in the recent year.

Topic modeling on the titles of four domains reveals the hot topics in each field. There are a total of 96, 41, 123, and 94 topics for data mining, software engineering, computer vision, and human-computer interaction, respectively. Here, we present the

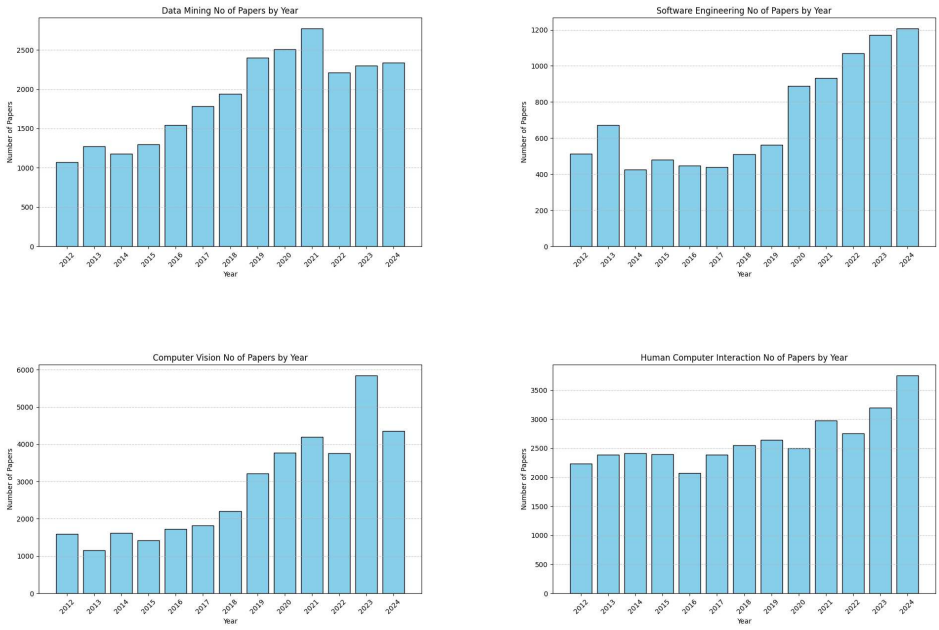


Fig 1. Yearly number of paper published from 2012 to 2024

top 10 topics for each domain in Tables 2, 3, 4, and 5, respectively. Topic -1 consists of articles that could not be clustered in any of the other groups. Each article is assigned to only one cluster.

Topic	Count	Name	Percentage
-1	9019	-1_inform retriev_neural network_ai_dataset	36.62%
0	1700	0_recommend graph_network recommend_sequenti recommend_recommend social	6.9%
1	1649	1_graph neural_graph convolut_graph learn_deep graph	6.7%
2	981	2_social data_social network_detect social_twitter data	3.98%
3	823	3_electron health_medic data_healthcar data_health record	3.34%
4	618	4_search evalu_search use_search queri_convers search	2.51%
5	540	5_learn rank_rank learn_rank model_rank bias	2.19%
6	319	6_anomali detect_detect anomali_unsupervis anomali_model anomali	1.3%
7	301	7_languag model_larg languag_spoken languag_understand languag	1.22%
8	297	8_cluster subspac_subspac cluster_cluster hypothesi_view cluster	1.21%
9	249	9_infer network_network infer_traffic detect_commun detect	1.01%

Table 2. Top 10 topics name in Data Mining and prominent articles within 2012 to 2024

From Table 2, we observe significant research activity in cutting-edge AI-driven topics which is visible in 20% of the papers. Graph learning and recommendation lead with 1700 papers and 1649 papers, folowed by social network (981 papers).

These topics indicate a strong research focus on structured knowledge representation, conversational AI, social media mining, and personalized information retrieval. Additionally, AI's rapid integration into Data Mining is evident, as topics such as large language model, anomaly detection, raking system are among top 10 topics. This shows that in the last decade there is a growing awareness and emphasis on AI deployment in data-driven applications.

In Software Engineering Table3, we can see that 12.27% of articles focus on haptic touch and 6.31% on software development, which indicates that robotics touch input and motion-based interfaces are becoming popular. We also observe that LLM is widely

Topic	Count	Name	Percentage
-1	3617	-1_repair_automat_code_program	38.84%
0	1143	0_haptic_touch_interact_hand	12.27%
1	588	1_softwar develop_softwar engin_softwar evolut_softwar architectur	6.31%
2	347	2_android test_test android_android applic_analysi android	3.73%
3	270	3_convolut neural_deep neural_robust deep_deep learn	2.9%
4	246	4_languag model_model languag_languag code_larg languag	2.64%
5	230	5_test javascript_analysi javascript_web test_web autom	2.47%
6	183	6_autom repair_automat repair_program repair_repair autom	1.97%
7	150	7_mobil app_mobil applic_interact mobil_develop mobil	1.61%
8	136	8_vulner detect_vulner neural_vulner predict_vulner dataset	1.46%
9	134	9_code review_code summar_code exampl_program analysi	1.44%

Table 3. Top 10 topics in Software Engineering and prominent articles within 2012 to 2024

utilized, accounting for 2.64% of papers. Other subjects include Android testing, deep learning, code summarization, vulnerability detection, and mobile applications.

Topic	Count	Name	Percentage
-1	12634	-1_object detect_neural network_vision_convolut	34.44%
0	1172	0_recognit facial_face recognit_recognit face_face reconstruct	3.20%
1	1114	1_segment video_video object_object segment_video segment	3.04%
2	870	2_learn pose_pose transfer_pose track_shape pose	2.37%
3	822	3_medic imag_imag segment_patholog imag_transform medic	2.24%
4	631	4_super resolut_super resolv_resolut network_resolut imag	1.72%
5	589	5_point cloud_pointnet_cloud point_learn point	1.61%
6	586	6_action recognit_recognit video_recognit action_action detect	1.60%
7	546	7_object track_track multi_visual track_object tracker	1.49%
8	543	8_person identif_identif person_person id_identif unsupervis	1.48%
9	534	9_adversari attack_attack adversari_adversari robust_adversari perturb	1.46%

Table 4. Top 10 topics in Computer Vision and prominent articles within 2012 to 2024

Topic	Count	Name	Percentage
-1	14789	-1_human robot_robot interact_motion plan_robot	43.15%
0	1903	0_wing uav_quadrotor uav_aerial vehicl_aerial robot	5.55%
1	869	1_surgeri robot_robot surgeri_invas robot_robot endoscop	2.54%
2	788	2_robot path_path plan_plan manipul_robot manipul	2.30%
3	754	3_predict autonom_trajectori predict_vehicl trajectori_autonom drive	2.20%
4	708	4_magnet microrobot_magnet actuat_control magnet_perman magnet	2.07%
5	672	5_robot team_multi robot_robot multi_multirobot	1.96%
6	668	6_exoskeleton control_limb exoskeleton_exoskeleton robot_exoskeleton assist	1.95%
7	599	7_robot grasp_grasp detect_grasp object_object grasp	1.75%
8	468	8_motion plan_sampl strategi_path plan_sampl base	1.37%
9	392	9_visual slam_slam visual_rgb slam_monocular slam	1.14%

Table 5. Top 10 topics in Human Computer Interaction and prominent articles within 2012 to 2024

In the Computer Vision Table 4, topics including face recognition, video segmentation, human pose tracking via transformers, medical imaging transformation, and super-resolution imaging represent the top 10% of all topics. Other areas, such as point clouds, person identification, object tracking, and adversarial attacks, are also notable.

In Human-Robot Interaction Table 5, the leading research topics are flapping wing, aerial and surgical robots and robot path planning that emphasizing the critical role of navigation and movement strategies in robotics. Lastly, microrobot, multi robot, grasp Detection and visual slam signifies a strong focus on robotic manipulation and dexterity, essential for applications in automation and assistive robotics.

We focused on investigating the emergence of AI in these areas. By employing topic modeling, we pinpointed AI-related subjects within each domain, such as graph neural networks and LLM in DM, as well as LLM in SW, deep learning in CV and reinforcement learning in HRI. These findings highlight the increasing integration of AI within these fields.

3.1 Topic Shifts

After determining the topics of each discipline, we examined the trends and growing interest in these subjects. We noticed that authors switched their published paper topics almost every year. Each topic has a number of authors moving into it, as well as a number of authors transitioning to another topic. Based on this, we calculated the topic shift. The topic shift for a given subject in a year is the difference between the number of authors published under that topic and the number of authors who moved out from that topic from the previous year to the current year. From this, we found that some topics show significant interest from authors in a given year, while others have seen a decline in interest over the years. When there are more authors moving than moving out, the shift value is positive, otherwise negative. In the year a topic emerged, the topic shift value is positive because the authors have recently published papers on that topic. In subsequent years, if more authors leave the topic than enter, then the value becomes negative. We start our analysis from 2012; if many authors leave the topic, it begins with a higher negative value. For the latest current year, since there is no information about authors leaving the topic after 2024, the value is positive. In data mining 2, we can that, topic 0_graph neural_deep graph_graph convolut_graph and topic 1_base recommend_network recommend_learn recommend_recommend social learn had increased interest from 2022 to 2023 and 2024. Topic 2_social data_social network_detect social_twitter data had a steady inclination until 2020, after that the growth has declined. The top 10 topics have comparatively more authors coming into the topic than moving out from it and that's what it gave the topic a steady growth. The topics which did not experience much growth because authors move in and out significantly every year, which is why it did not reach a major important topic.

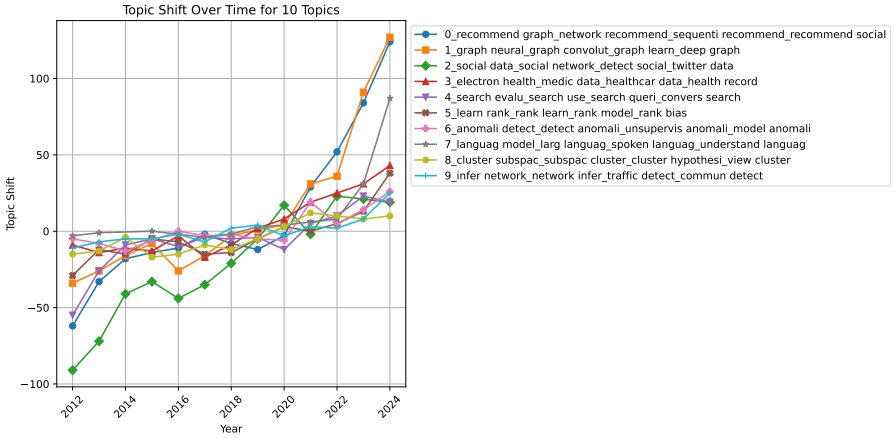


Fig 2. Data Mining Topics Shift and Trends

Similarly, in Software Engineering, in top 10 3, topic 4_languag model_model languag_languag code_larg languag has the highest growth from 2022 to 2024 though it had a stead progress since 2012. Additionally, we can observe that 1_softwar develop_softwar engin_softwar evolut_softwar architectur interest have experienced fluctuations over the past decade.

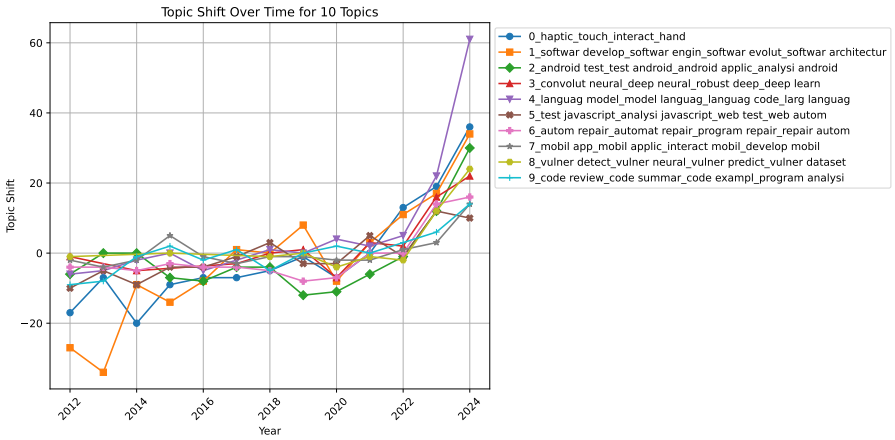


Fig 3. Software Engineering Topics Shift and Trends

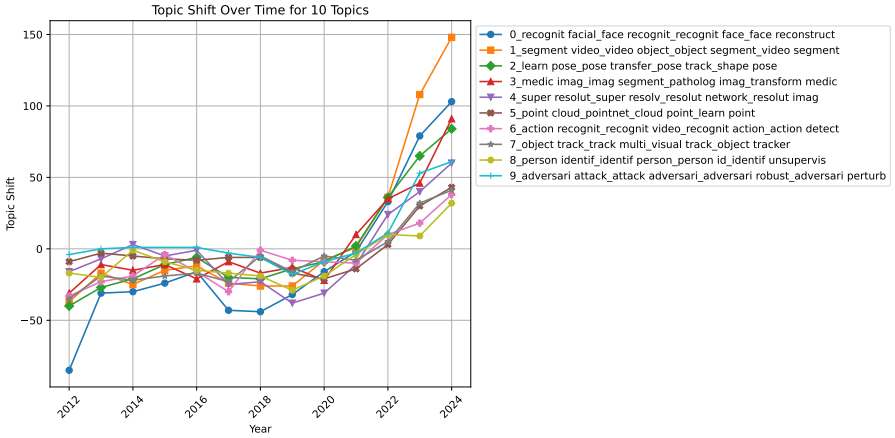


Fig 4. Computer Vision Topics Shift and Trends

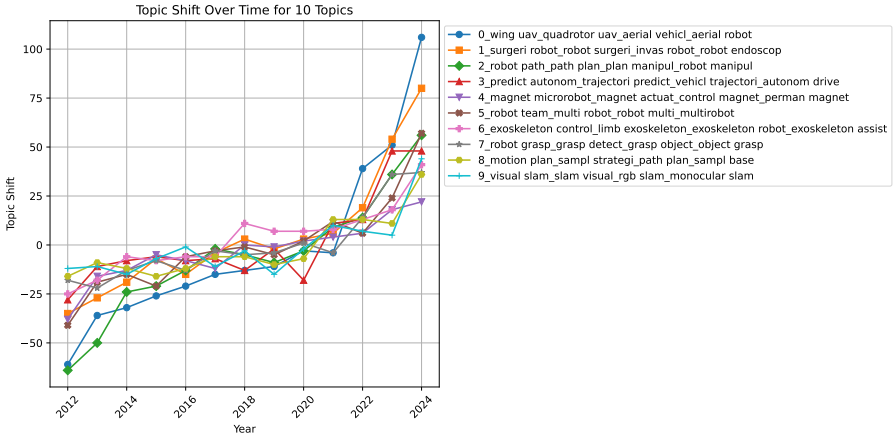


Fig 5. Human Computer Interaction Topics Shift and Trends

In Computer Vision, 1_segment video_video object_object segment_video segment has higher grownth interest. Most topics experienced slow growth until 2020. In 2016, for example, it is evident that all topics shifts were located very close to each other. It is

possible that the authors were moving in and out within these topics or shifting to other trending subjects that year. Similarly, in HRI, all the topics show steady upward growth. Topic 0_wing uav_quadrotor uav_aerial vehicl_aerial began with the lowest downturn and reached its highest peak within a decade.

The accuracy of topic shift analysis relies on improved topic modeling. The better articles are clustered, the more promising the results for topic shifts. Some papers may pertain to multiple topics. Analyzing overlapping topics in papers might provide a better overview of topic shifts.

4 Constructing co-authorship networks

Co-authorship networks are extensively used to understand scientific collaboration and the position of individual authors in the network. In co-authorship networks, authors know each other and it implies temporal and collegial relationship among authors all over the world. Many previous work has been analyzed on the co-authorship network [34–37]. Our purpose of analyzing co-authorship network is to compare different research disciplines and understand cross-cooperation among them.

We filter the data set within the time frame of 2022 to 2025 to build an undirected and weighted network using NetworkX. After selecting publication venues during this period, we constructed an attributed graph, where each node has a name, ORCID ID, and affiliation information. An edge between two authors is created if they have co-authored an article and the weight value of an edge is the year they have worked together. For example, if an article has three authors, each pair of co-authors will be connected by an edge, forming a collaborative structure that reflects research partnerships. Table 6 presents the dataset information.

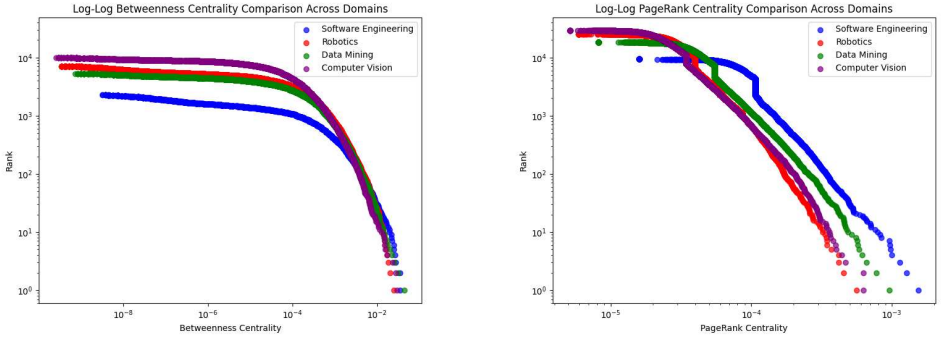
Research Domain	#Papers	#Authors	#Connections	#Publication Venue	Density
Data Mining	6,851	19,623	81,269	5	0.00045
Software Engineering	3,448	9,417	32,285	6	0.00072
Computer Vision	13,962	29,151	142,277	4	0.00033
Human Robot Interaction	9,710	25,239	124,056	5	0.00038

Table 6. Basic Information of DBLP Research Domain Networks within 2022 to 2025

To understand the overall distribution of nodes in the network, we plotted betweenness centrality of nodes. It shows how nodes connect with each other and allow for comparison of network structure in terms of node connectedness. Betweenness of node is defined as the number of shortest paths between all possible pairs of nodes traverse through the node. [38]. Figure 6(a) illustrates the comparison of betweenness centrality on a logarithmic scale across four disciplines. The X-axis represents the value of betweenness centrality in logarithmic scale, while the Y-axis displays the rank of authors in logarithmic scale, where lower rank values (at the top of the plot) correspond to authors with high betweenness, and higher rank values (at the bottom of the plot) correspond to authors with lower betweenness centrality. Here, Software Engineering shows a distinct pattern where authors have lower betweenness centrality compared to other domains. Robotics has a slightly higher centrality than Data Mining, indicating more influential nodes in Robotics compared to Data Mining, while Computer Vision has the highest betweenness centrality, suggesting that top-ranked authors are often well-connected in collaborations.

PageRank [39] is a technique for identifying relatively important nodes in a network. Figure 6(b) illustrates a comparison of PageRank centrality on a log scale for four disciplines. The values on the X-axis indicate the more influential authors in the research network, while the Y-axis (in the top plot, lower values) corresponds to highly ranked influential authors, and higher values (at the bottom) represent less central

authors. The log-log nature of the graph suggests that PageRank centrality follows a power-law distribution, indicating that a small number of authors have very high influence, while most have low influence. The distribution of authors in Software Engineering appears as a more scattered curve. Computer Vision and Robotics host the most influential authors, implying strong collaboration. Data Mining researchers fall between Robotics and Software Engineering, suggesting they have a moderate level of connectivity.



(a) **(b)**

Fig 6. (a)Betweenness Centrality and (b)PageRank comparison of four disciplines

4.1 Unveiling Cross-Disciplinary Collaborations in Computer Science

In this section, we explore collaboration across disciplines and identify the key players who contribute to various fields within past three years. To achieve this, we searched for authors who have published papers in three research domains. Due to ambiguity, we only considered the names of authors with unique ORCID IDs or DBLP numeric identifiers. By identifying and analyzing authors who contribute to multiple domains, we aim to uncover the interconnectedness of research efforts and the increasing convergence of methodologies and challenges across these areas.

The growing complexity and scope of modern research have led to the emergence of multidisciplinary approaches that span various domains. By examining the overlapping authors, we can identify key researchers who bridge gaps between different fields and promote cross-domain collaborations. This analysis helps us understanding how ideas, techniques, and innovations are shared across disciplines, driving advancements that may not be achievable within the constraints of a single research area.

Across the three domains of Data Mining, Computer Vision, and Human-Robot Interaction (HRI), we identified 139 unique authors who have contributed to all three fields and 408 connections. Similarly, in the combination of Computer Vision, Software Engineering, and HRI, there are 95 authors and 285 connections with contributions spanning these areas. The intersection of Software Engineering, Computer Vision, and Data Mining includes 71 authors with 204 connections who have actively worked across these domains. Lastly, Data Mining, Software Engineering, and HRI have a total of 31 authors and 84 connections contributing to all three research areas. Because we have only considered unique authors, there can be some more authors who have contributed, but we could not distinguish uniquely across the domains because of ambiguity.

Figure 7 illustrates the bipartite subgraphs of three disciplinary collaboration. In these subgraphs, the weight of the edges between the discipline node and author node

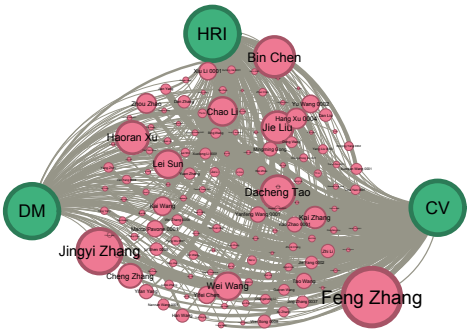
represents the number of papers published within that domain, and the size of the author node reflects the average degree value from all three disciplines for that author.

Usually, we observe that Data Mining, Computer Vision, and Human-Robot Interaction (HRI) share foundational concepts, particularly in machine learning, reinforcement learning, pattern recognition, and data-driven modeling, which facilitate greater interdisciplinary collaborations. This is reflected in our findings 7a, where 139 unique authors and 408 connection among these three.

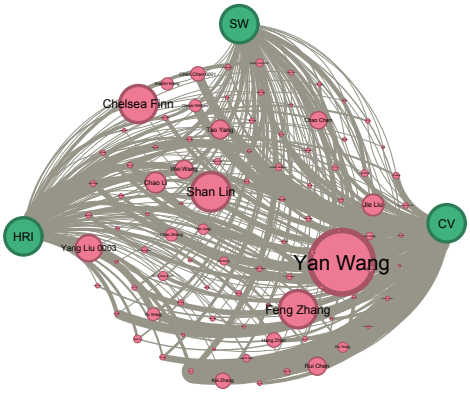
For example, one researcher "Xiu Li 0001" has contributed to all three fields, highlighting the strong interdisciplinary nature of research. The research contributions of this author span multiple disciplines, showcasing significant cross-domain collaboration. His work in HRI [40–42] focuses on robotic perception, locomotion, and control strategies for various applications, such as underwater imaging, micromanipulation, and exoskeletons. Simultaneously, his research in Data Mining [43–45] demonstrates expertise in recommender systems, query classification, and predictive modeling, emphasizing hierarchical learning and contrastive learning techniques. Additionally, his contributions to Computer Vision [46–48] explore topics such as video comprehension, diffusion models, camouflaged object detection, and 3D choreography generation, reflecting a strong foundation in deep learning and image analysis. The breadth of his research across these fields highlights their ability to integrate machine learning techniques across different domains, driving interdisciplinary innovation at the intersection of robotics, data-driven decision-making, and visual intelligence. This explanation emphasizes authors ability to bridge gaps between fields and apply machine learning and data-driven approaches in robotics, recommender systems, and computer vision, making them a key contributor to interdisciplinary research.

Similarly, Computer Vision, Software Engineering, and Human Computer Interaction Fifure 7b exhibit 95 shared contributors and 285 connections, suggesting an overlap in areas such as robotic perception, vision and control, and software development for intelligent interfaces. For example, Achuta Kadambi's research contributions span across Software Engineering, Computer Vision, and Human-Robot Interaction (HRI), demonstrating a strong interdisciplinary approach. In Software Engineering, the author's work on ForceSight: Non-Contact Force Sensing with Laser Speckle Imaging [49] explores innovative sensing techniques, which have implications for real-world applications such as robotics and healthcare; in Computer Vision research covers a diverse range of topics, including single-image deweathering (WeatherStream) [50], 3D Gaussian [51] splatting for feature fields (Feature 3DGS), depth estimation using multi-modal sensor fusion (Depth Estimation from Camera Image and mmWave Radar Point Cloud) [52], and polarimetric coordinate networks for neural scene representations (pCON) [53] and in HRI, the author's contribution to All-day Depth Completion further demonstrates [54] shows the ability to develop vision-based solutions for autonomous systems and robotics. The author's research bridges multiple domains, integrating advanced sensing, computer vision, and machine learning to tackle complex challenges in perception, automation, and human-robot collaboration.

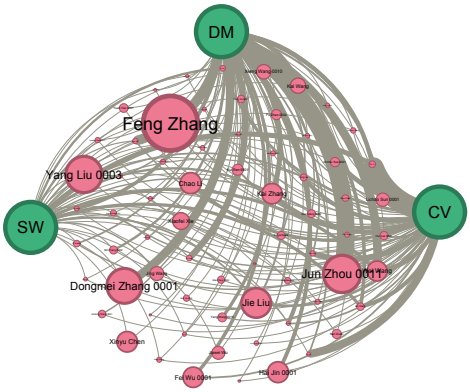
The intersection of Software Engineering, Computer Vision, and Data Mining 7c has 71 authors and 204 connections, indicates the role of scalable software frameworks, algorithm optimization, and data-driven decision-making in both engineering and vision-based applications. For example, Nan Dans' work in Computer Vision, focuses on vision-language models and generative AI, as seen in ReCo: Region-Controlled Text-to-Image Generation [55], which enhances controllability in text-to-image synthesis, and VL-InterpreT [56], which provides interactive visualization tools for interpreting vision-language transformers. The author's work in video analysis is reflected in



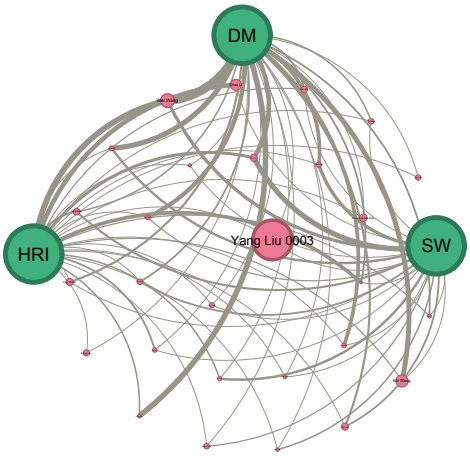
(a) Key players across Data Mining, Human Robotics Interaction and Computer Vision



(b) Key players across Computer Vision, Human Robotics Interaction and Software Engineering



(c) Key players across Computer Vision, Data Mining and Software Engineering



(d) Software Engineering, Data Mining and Human Robot Interaction

Fig 7. Key players in interdisciplinary research

Learning Temporal Video Procedure Segmentation, which focuses on structuring large-scale video datasets. In Software Engineering, the author contributes to automated programming and AI-driven software development. Contextualized Data-Wrangling Code Generation in Computational Notebooks explores AI-assisted code generation [57], improving automation in data science workflows, while Automating Code Review Activities by Large-Scale Pre-Training applies deep learning to streamline software review processes. In Data Mining, the author's research focuses on information retrieval and knowledge distillation. PROD: Progressive Distillation for Dense Retrieval [58] and LEAD: Liberal Feature-Based Distillation for Dense Retrieval [59] contribute to improving large-scale information retrieval through advanced distillation techniques.

Nan Duan's research trajectory shows gradual expansion from structured data mining and retrieval systems toward generative AI, vision-language models, and AI-assisted software engineering. This shift follows a broader trend in academia and industry, where researchers increasingly integrate natural language processing, vision, and retrieval models to develop more interactive, AI-powered systems. It also reflects the convergence of AI techniques across domains, as large-scale pre-training and multimodal learning become central to advancements in software automation, content generation, and information retrieval.

Meanwhile, Data Mining, Software Engineering, and Human Computer Interaction 7d have a relatively smaller overlap of 31 authors and 84 connections, possibly due to more distinct methodologies in these fields, though they still intersect in areas like intelligent automation, user-driven analytics, and adaptive learning systems. One example of author Pingchuan Ma's research in Software Engineering shows [60, 61] how causal analysis and chaos engineering can enhance software performance and fairness in machine learning models. In Data Mining, his papers [62, 63] explore innovative methods for causal inference, offering scalable solutions to complex problems in data-driven decision-making. In Human-Robot Interaction (HRI), Ma's paper "ASAP: Automated Sequence Planning for Complex Robotic Assembly with Physical Feasibility" [64] stands out, applying his expertise in causality and sequence planning to solve real-world problems in robotics. His contributions reflect the growing trend of interdiscipline collaboration, where advances in one field often inform and enhance methodologies in others. We have also tried to find if there is any author participated in all four disciplines. But we could not identify any author who is common in all four.

By recognizing these overlaps, we pinpoint areas where interdisciplinary research is gaining momentum, providing new directions and opportunities for future work. It not only generates fresh ideas but also expands the knowledge of computer science in every direction.

5 Institutional Collaboration within Disciplines

5.1 Institutions Count

Researchers not only showcase their ideas and work in published articles, but also the affiliated institutions they represent play a vital role in these publications. For example, the topics researchers pursue are clearly aligned with the interests of their institutions and the research areas where these institutions are willing to invest. Therefore, apart from the authors' interests, we are also eager to see which institutions are actually involved in this work. To achieve this, we have listed the institutions to which the researchers belong for each discipline. We classified the institutions into four categories: University, Research Centers (governmental or private research centers), Company, and Company Research Lab (owned and run by companies).

We counted the total number of authors affiliated with institutions. For this

purpose, we considered the unique authors and their affiliation information. We did not consider authors whose affiliation information is not available in their ORCID profile. For institution type categorization, we used the OpenAI (GPT-4 Turbo) [65] Large Language model. In all four disciplines, a total of 5,356 unique institutions were identified, and we manually verified more than 200 institutions that they were properly categorized by the LLM. Table 7 shows the top 20 highest count institutions that appeared across all four disciplines.

Institution	Count	Category
Alibaba Group	134	Company
Tsinghua University	123	University
Zhejiang University	114	University
Amazon	108	Company
Shanghai Jiao Tong University	72	University
Fudan University	70	University
Tencent	68	Company
University of Science and Technology of China	67	University
Peking University	64	University
Meta	63	Company
Beihang University	59	University
Beijing University of Posts and Telecommunications	59	University
Huawei Technologies	57	Company
LinkedIn	55	Company
Nanyang Technological University	54	University
Google	50	Company
East China Normal University	49	University
City University of Hong Kong	47	University
Microsoft	46	Company
Hong Kong University of Science and Technology	44	University

Institution	Count	Category
Zhejiang University	65	University
Carnegie Mellon University	47	University
Nanjing University	36	University
Tsinghua University	36	University
Fudan University	35	University
Alibaba Group	31	Company
Peking University	31	University
Meta	28	Company
Sun Yat-sen University	27	University
Google	25	Company
Shanghai Jiao Tong University	25	University
City University of Hong Kong	22	University
University of Washington	21	University
National University of Defense Technology	21	University
Huawei Technologies	21	Company
Nanyang Technological University	20	University
Hong Kong University of Science and Technology	20	University
Aarhus University	18	University
Microsoft Research	18	Company Research Lab
Beijing University of Posts and Telecommunications	18	University

Institution	Count	Category
Tsinghua University	118	University
Zhejiang University	104	University
Shanghai Jiao Tong University	90	University
Peking University	83	University
Alibaba Group	82	Company
City University of Hong Kong	81	University
Nanyang Technological University	70	University
Fudan University	64	University
Tencent	63	Company
University of Science and Technology of China	54	University
Meta	53	Company
University of Electronic Science and Technology of China	50	University
National University of Singapore	49	University
Huazhong University of Science and Technology	48	University
Google	48	Company
Carnegie Mellon University	47	University
Beijing University of Posts and Telecommunications	47	University
Huawei Technologies	46	Company
Hong Kong University of Science and Technology	45	University
Xi'an Jiaotong University	42	University

Institution	Count	Category
Tsinghua University	75	University
Zhejiang University	63	University
Carnegie Mellon University	55	University
KTH Royal Institute of Technology	54	University
University of Michigan	50	University
City University of Hong Kong	48	University
Shanghai Jiao Tong University	45	University
Nanyang Technological University	45	University
Delft University of Technology	44	University
ETH Zurich	41	University
Istituto Italiano di Tecnologia	40	Research Center
Technical University of Munich	39	University
Peking University	38	University
Massachusetts Institute of Technology	37	University
Alibaba Group	36	Company
Imperial College London	33	University
Fudan University	33	University
University of Science and Technology of China	32	University
National Institute of Advanced Industrial Science and Technology	32	Research Center
University of Electronic Science and Technology of China	31	University

Table 7. Comparison of Affiliated Institution Counts of authors (top left: Data Mining), (top-right: Software Engineering, (bottom-left: Computer Vision), (bottom-right: Human Computer Interaction)

From the table, we can also see that Data Mining has 7 companies, Software Engineering has 4 companies, Computer Vision has 5, and HRI does not have a company affiliation in the top 20. Most of them are affiliated with universities in the top 20, while Software Engineering includes Microsoft Research Lab, and HRI has two research centers involved. We distinctly categorized governmental or nongovernmental research labs and company labs because company research labs might be more biased by company interests and profits. Additionally, there are some institutions that cannot be categorized by LLM; in those cases, we categorize them with others. The distribution of each category in the total dataset showed in Figure 8

The distribution shows that academic institutions dominate research across all four fields, highlighting the strong role of universities in advancing knowledge. Industry involvement is highest in Data Mining and Computer Vision, likely due to their strong commercial applications, while Software Engineering also sees notable company contributions, though lower than Data Mining. Human-Robot Interaction (HRI) has the least company participation but stands out for having more contributions from Research Centers compared to other fields, possibly indicating a greater focus on experimental and interdisciplinary research. Additionally, Research Labs from companies show

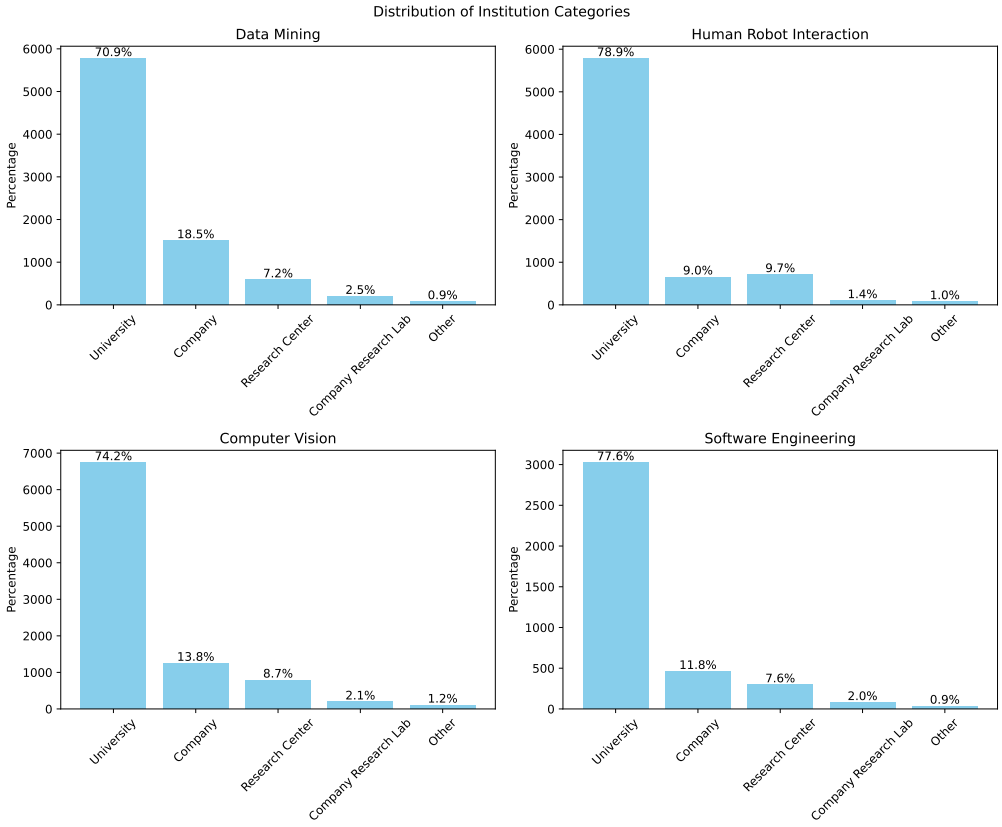


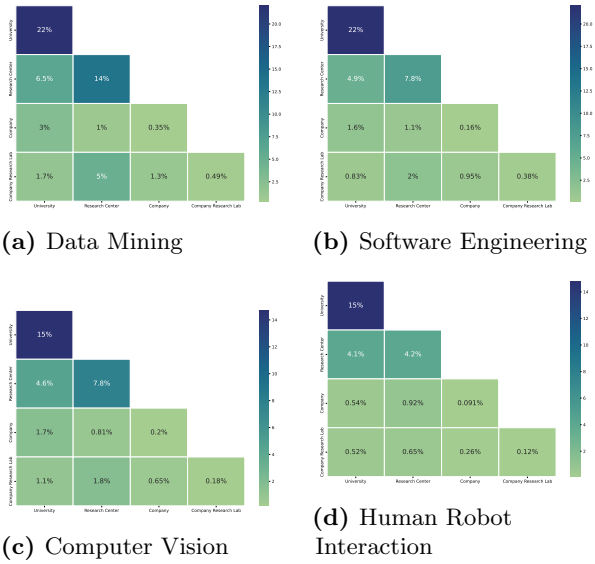
Fig 8. Distribution of type of institutions across domains

significant interest in Data Mining but relatively lower involvement in HRI, suggesting that industries prioritize fields with immediate business applications over those requiring longer-term development.

This information leads to the next question: how and to what extent do these affiliated institutions collaborate among themselves. The following sections describe the collaboration across institutions within disciplines.

5.2 Collaboration among Institutions Type

According to Philip S. Morrison et al. [66], collaboration is heavily influenced by both intellectual and physical proximity. Close physical proximity in sharing ideas remains crucial at both institutional and regional levels. How universities organize their research environments, both organizationally and spatially, can significantly affect research productivity. Universities tend to have tightly-knit groups within or between departments that collaborate more effectively than other institutions. Professors and students, from undergraduate to post-doctoral levels, all contribute significantly to research in academic institutions. Although funding opportunities vary from one university to another, the environment facilitate most of the research occurring in universities, as reflected in our analysis outcomes. Our analysis indicates that across all disciplines, universities are significantly more productive in terms of research, publications, and collaboration. From Figure 9, we can observe that most collaborations occur among universities. The value of each block in the heatmap represents the percentage of authors involved in this type of collaboration out of the total number of authors in that domain. There are many authors for whom we do not know affiliation



(a) Data Mining **(b)** Software Engineering

(c) Computer Vision **(d)** Human Robot Interaction

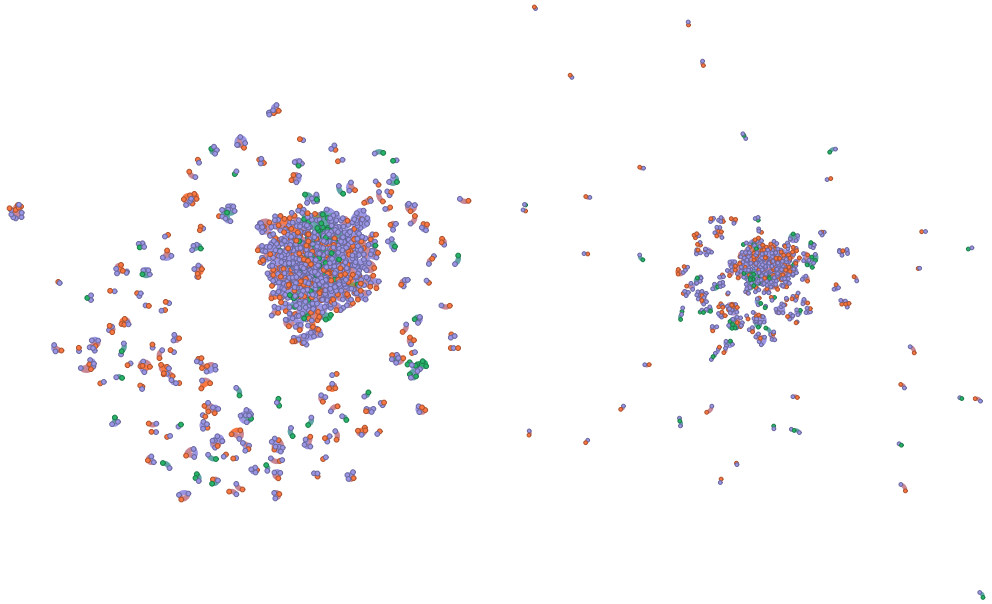
Fig 9. Percentage of Institutions(nodes) in Different types of Collaboration

information. We observe that in Data Mining and Software Engineering, the highest collaboration types exists among universities, at 22% percent, while in Computer Vision and Robotics have 15%. Research Centers also collaborate with universities, primarily in Data Mining at 6.5% and other domains around 4-5% in general, though Research Centers are work more among or within themselves. Companies show more interest in collaborating with universities than with one another. For Data Mining, collaboration between research centers and universities is at 3%; for Software Engineering, it is 1.6% percent; for Computer Vision, it is 1.7%; and for Robotics, it is 0.54%. Conversely, collaboration between companies is significantly lower, at 0.35%, 0.16%, 0.2%, and 0.091% respectively. Even company research labs are more inclined to partner with universities compared to collaborating internally on research.

This analysis highlights the dominant role of universities in research collaboration and productivity across all disciplines. The strong collaboration within and between universities suggests that intellectual and environment fosters research output. Research Centers also show substantial collaboration with universities, particularly in Data Mining, which may indicate a shared focus on applied research. Interestingly, companies demonstrate a preference for collaborating with universities rather than with other companies, likely due to access to academic expertise, fresh talent, and innovative research. The significantly lower company-to-company collaboration rates suggest that competitive interests and proprietary concerns may hinder direct industry partnerships. Even company research labs seem to favor academic collaborations over internal research, reinforcing the idea that universities serve as key innovation partners for both private and public research entities.

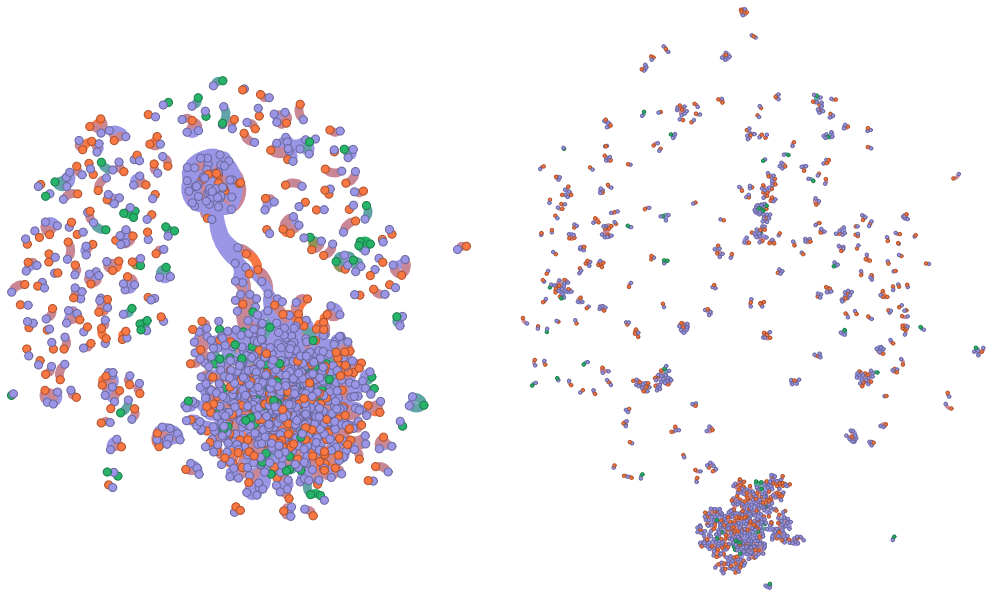
5.3 Collaboration Subgraphs

We extracted three types of collaboration subgraphs from the original graphs to visualize the high-level structure of University-Research Centers, University-Company, and Company and Research Centers collaboration. Here, for subgraph, we combined governmental, non-governmental research centers and company research centers together. Figure 10 illustrates the four discipline subgraphs on University-Research Center Collaboration. Figure 11 illustrates the four disciplines subgraphs on



(a) Data Mining

(b) Software Engineering



(c) Computer Vision

(d) Human Robot Interaction

Fig 10. Collaboration between University and Research lab affiliated authors (Purple (University), Orange (Research Center), and Green (Company Research Lab))



(a) Data Mining

(b) Software Engineering

(c) Computer Vision

(d) Human Robot Interaction

Fig 11. Collaboration between University and Company affiliated authors (Blue: Company, Purple: University)



(a) Data Mining **(b)** Software Engineering

(c) Computer Vision **(d)** Human Robot Interaction

Fig 12. Collaboration between Company and Research Center affiliated authors (Blue: Company, Orange: Research Center, Green: Company Research Lab)

Company-University Collaboration overviews. Additionally, Figure 12 illustrates the four disciplines of Company-Research Center collaboration. In this illustration, nodes are represented in five colors colors: purple (university), blue (company), orange (research center), and green (company research lab).

In all figures, we observe the largest connected component, where nodes are closely interconnected. There are many small clusters.

5.3.1 University-Research Center Collaboration

As most research happens in universities, it is interesting to investigate how research centers contribute to it. From the subgraphs 10, we can see that Data Mining and Computer Vision have more tightly knit groups compared to Software Engineering and Robotics. In Robotics, there are more small groups than in Software Engineering, along with one large group.

In Data Mining, the top institutions contributing are Zhejiang University, Microsoft Research Asia, Tsinghua University, the University of Science and Technology of China, and Microsoft Research. In Software Engineering, the top institutions are Zhejiang University, Microsoft Research Asia, Microsoft Research, CSIRO, and the University of Luxembourg.

The top institutions contributing to Computer Vision are Tsinghua University, Zhejiang University, Shanghai Jiao Tong University, Peking University, and Nanyang Technological University, which are mostly universities. In the field of Computer Vision, we see two connected clusters, with several other nodes linking them, thereby forming a smaller component. Upon investigating the authors within this smaller component, we discovered that the cluster was formed due to the paper "Why is the winner the best?".

In Robotics, the top institutions are Istituto Italiano di Tecnologia, the National Institute of Advanced Industrial Science, KTH Royal Institute of Technology, Zhejiang University, and Nanyang Technological University.

From above discussion, we can say that in Data Mining and Software Engineering, Microsoft Research and Microsoft Research Asia play a significant role, showing strong corporate involvement in these fields. In Computer Vision and Robotics, the top contributors are primarily universities, suggesting that industry involvement in these fields may be more distributed or focused on specific areas. Also, Chinese universities (Zhejiang, Tsinghua, Shanghai Jiao Tong, Peking, and USTC) have a strong presence across multiple fields, highlighting China's growing influence in AI and related domains. European institutions, such as the Istituto Italiano di Tecnologia and KTH, play a significant role in Robotics, suggesting regional expertise in this domain.

5.3.2 University-Company Collaboration

In University Company Collaboration Figure 11, we can learn which companies are more interested in collaborating with universities. From the subgraphs analysis, we have seen that Data Mining and Computer Vision have large connected groups, while Software Engineering has scattered groups with midsize connected clusters. In Data Mining, the top institutions that have contributed are Alibaba Group, Amazon, Tencent, Tsinghua University, and Zhejiang University. In Software Engineering, the top contributors are Zhejiang University, Meta, Google, Tencent, and Alibaba Group. In Computer Vision, the leading institutions are Tencent, Alibaba Group, Meta, Zhejiang University, and Google. In Robotics, the principal contributors include Tencent, Tsinghua University, State Grid Corporation, Technical University of Munich, and Alibaba Group.

From the analysis, we can say that Amazon is focused on data mining, Google and Meta have an interest in software engineering and computer vision, and Alibaba Group is interested in all disciplines while collaborating with universities. These patterns

reflect how different companies strategically engage with academia based on their research priorities.

5.3.3 Company-Research Center Collaboration

The research center includes both company labs and government-affiliated research centers. From these subgraphs, we can learn the companies are interested in researching with their own research centers or other research centers 12. In terms of data mining, the majority of collaborations come from Microsoft Research Asia, Alibaba Group, Microsoft, Tencent, Microsoft Research, and Meta. As shown in Figure 12a, there is a large connected group, with Microsoft Research Asia showing the most involvement. Additionally, there are several scattered contributions, along with a dedicated group from Meta in this area. In Software Engineering 12b, the top five collaborators are Microsoft Research Asia, Microsoft Research, Microsoft Company, Meta, and Google. It's interesting that, although the number of publications is less compared to the other three domains, the big tech companies and research centers are actively involved in publishing papers here. In the largest connected group, we again see that Microsoft Research Asia is involved. In the field of Computer Vision, the leading intitutions are Google, Tencent, Microsoft Research Asia, Microsoft Research, and Alibaba Group. Among these, Google and Microsoft Research form the largest connected group and have the most collaborations. In Robotics, the top organizations include the Italian Institute of Technology, Microsoft Research, Honda Research Institute, Tencent, and Nvidia. The graph shows that there is one dedicated team from the Honda Research Institute and one dedicated group from Microsoft Research. Additionally, Nvidia is actively involved in the largest connected group.

Based on these observations, we can say that Microsoft has an interest in publishing papers across all four domains collaborating with Research Center or within their lab to company, while other companies like Google, Meta, Tencent, and Nvidia show selective interest in research.

Discussion

Analyzing scientific collaboration across different subdisciplines helps to identify potential growth of a domain. To effectively explore collaboration, subdisciplines should be examined both individually and in relation to one another. Computer Science was chosen as an exemplary case to demonstrate an approach for examining a domain from different perspectives. While our approach provides valuable insights into cross-disciplinary collaboration in Computer Science, it has several limitations. We rely on specific databases, such as DBLP and ORCID, as well as selective conferences, which may not reflect the entire landscape of computer science. Incorporating more disciplines offers a more comprehensive view of the current state. However, our approach outlines the necessary steps for making such comparisons. This approach can be applicable to any other research domains too, such as studying science. Topic modeling provides an overview of explored topics, while trend analysis reveals emerging research areas. By understanding these topics, collaboration patterns can be better analyzed—identifying which domains share similar knowledge and where opportunities for new collaborations exist. If no prior collaboration has occurred between certain domains, it can indicate if there is potential for interdisciplinary research or not. Furthermore, from research to transition into commercial applications, industry collaboration gives information for partnership.

It is hard to delve into what is happening in a field as an outsider. Based on demographics, research interest also varies. Through this analysis, the younger

generation who is interested in research can gain a comprehensive understanding of a field and be inspired to contribute to it. Researchers from one subdiscipline can also learn about another, broadening their knowledge and fostering interdisciplinary collaboration. More importantly, researchers from one domain can explore emerging trends of another domains, understand the research criteria and culture, and learn to incorporate that into their own field. For example, suppose a computer scientist is interested in collaborating with a biologist but lacks familiarity with the biological domain. Through this holistic approach, the computer scientist can develop a foundational understanding of biology, identify areas, and discover new opportunities for inter-domain research. Over time, collaborations can lead into groundbreaking innovations in both fields through shared knowledge and expertise. Not only because of personal interest, but also the collective interest of institutions can benefit with such kind of collaboration. The more institutions are interested in investing, the smaller the gap between domains becomes. From there, research can extend beyond academia into real-world impact which can give progress of mankind and ecological advancement. Lastly, as a future work plan, we may consider extending the proposed approach by using multi-dimensional information to recommending research partnerships with industries based on mutual interests and technological advancements.

Acknowledgments

We thank Dr. Erik Saule for being involved in study design, data collection, result analysis, reviewing the article, and providing valuable feedback.

References

1. Fontelo P, Liu F. A review of recent publication trends from top publishing countries. Systematic reviews. 2018;7:1–9.
2. Fanelli D, Larivière V. Researchers' individual publication rate has not increased in a century. PloS one. 2016;11(3):e0149504.
3. Vrettas G, Sanderson M. Conferences versus journals in computer science. Journal of the Association for Information Science and Technology. 2015;66(12):2674–2684.
4. Dupont PE, Nelson BJ, Goldfarb M, Hannaford B, Menciassi A, O'Malley MK, et al. A decade retrospective of medical robotics research from 2010 to 2020. Science robotics. 2021;6(60):eabi8017.
5. Yun JJ, Won D, Jeong E, Park K, Yang J, Park J. The relationship between technology, business model, and market in autonomous car and intelligent robot industries. Technological Forecasting and Social Change. 2016;103:142–155.
6. Ji YA, Nam SJ, Kim HG, Lee J, Lee SK. Research topics and trends in medical education by social network analysis. BMC medical education. 2018;18:1–10.
7. Fteimi N, Lehner F. Main research topics in knowledge management: A content analysis of ECKM publications. Electronic Journal of Knowledge Management. 2016;14(1):pp3–15.
8. Pan C, Li W. Research paper recommendation with topic analysis. In: 2010 International conference on computer design and applications. vol. 4. IEEE; 2010. p. V4–264.

9. Ghosh S, Chen Y, Dou W. Railroad safety: A systematic analysis of Twitter data. Case Studies on Transport Policy. 2024;15:101154.

10. Abdelrazek A, Eid Y, Gawish E, Medhat W, Hassan A. Topic modeling algorithms and applications: A survey. Information Systems. 2023;112:102131.

11. Shi T, Kang K, Choo J, Reddy CK. Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. In: Proceedings of the 2018 world wide web conference; 2018. p. 1105–1114.

12. Yan X, Guo J, Lan Y, Cheng X. A biterm topic model for short texts. In: Proceedings of the 22nd International Conference on World Wide Web. WWW '13. New York, NY, USA: Association for Computing Machinery; 2013. p. 1445–1456. Available from: https://doi.org/10.1145/2488388.2488514.

13. Moody CE. Mixing Dirichlet Topic Models and Word Embeddings to Make lda2vec; 2016. Available from: https://arxiv.org/abs/1605.02019.

14. Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space; 2013. Available from: https://arxiv.org/abs/1301.3781.

15. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding; 2019. Available from: https://arxiv.org/abs/1810.04805.

16. Abbasi A, Altmann J. On the correlation between research performance and social network analysis measures applied to research collaboration networks. In: 2011 44th Hawaii international conference on system sciences. IEEE; 2011. p. 1–10.

17. Li L, Catala-Lopez F, Alonso-Arroyo A, Tian J, Aleixandre-Benavent R, Pieper D, et al. The global research collaboration of network meta-analysis: a social network analysis. PloS one. 2016;11(9):e0163239.

18. Fagan J, Eddens KS, Dolly J, Vanderford NL, Weiss H, Levens JS. Assessing research collaboration through co-authorship network analysis. The journal of research administration. 2018;49(1):76.

19. Woo SH, Kang DJ, Martin S. Seaport research: An analysis of research collaboration using Social Network Analysis. Transport Reviews. 2013;33(4):460–475.

20. Zhang L, Qian Y, Ma C, Li J. Continued collaboration shortens the transition period of scientists who move to another institution. Scientometrics. 2023; p. 1–20. doi:https://doi.org/10.1007/s11192-022-04617-x.

21. Choe H, Lee DH. The structure and change of the research collaboration network in Korea (2000–2011): Network analysis of joint patents. Scientometrics. 2017;111:917–939.

22. Ibanez A, Bielza C, Larranaga P. Relationship among research collaboration, number of documents and number of citations: a case study in Spanish computer science production in 2000–2009. Scientometrics. 2013;95:689–716.

23. Franceschet M. Collaboration in computer science: A network science approach. Journal of the American Society for Information Science and Technology. 2011;62(10):1992–2012.

24. Bird C, Devanbu P, Barr E, Filkov V, Nash A, Su Z. Structure and dynamics of research collaboration in computer science. In: Proceedings of the 2009 SIAM International Conference on Data Mining. SIAM; 2009. p. 826–837.

25. Afolabi IT, Ayo A, Odetunmibi OA. Academic collaboration recommendation for computer science researchers using social network analysis. Wireless Personal Communications. 2021;121(1):487–501.

26. Pham MC, Klamma R, Jarke M. Development of computer science disciplines: a social network analysis approach. Social Network Analysis and Mining. 2011;1:321–340.

27. Huang J, Zhuang Z, Li J, Giles CL. Collaboration over time: characterizing and modeling network evolution. In: Proceedings of the 2008 international conference on web search and data mining; 2008. p. 107–116.

28. Lara-Cabrera R, Cotta C, Fernández-Leiva A. An analysis of the structure and evolution of the scientific collaboration network of computer intelligence in games. Physica A: statistical mechanics and its applications. 2014;395:523–536.

29. Palmer CL. Work at the boundaries of science: Information and the interdisciplinary research process. Springer Science & Business Media; 2013.

30. Ghosh S. Examining Different Research Communities: Authorship Network. In: Cherifi H, Donduran M, Rocha LM, Cherifi C, Varol O, editors. Complex Networks & Their Applications XIII. Cham: Springer Nature Switzerland; 2025. p. 71–82.

31. DBLP. DBLP XML Dataset; 2024. Available from: `https://dblp.org/xml/`.

32. ORCID. ORCID Public Data File 2024; 2024. Available from: `https://doi.org/10.23640/07243.27151305.v1`.

33. Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure; 2022. Available from: `http://arxiv.org/abs/2203.05794`.

34. Newman MEJ. Scientific collaboration networks. I. Network construction and fundamental results. Physical Review E. 2001;64(1):016131. doi:10.1103/PhysRevE.64.016131.

35. Smeaton AF, Keogh G, Gurrin C, McDonald K, Sødring T. Analysis of papers from twenty-five years of SIGIR conferences: what have we been doing for the last quarter of a century? ACM SIGIR Forum. 2002;36(2):39–43. doi:10.1145/792550.792556.

36. Chen C, Carr L. Trailblazing the literature of hypertext: author co-citation analysis (1989–1998). In: Proceedings of the tenth ACM Conference on Hypertext and hypermedia : returning to our diverse roots: returning to our diverse roots. Darmstadt Germany: ACM; 1999. p. 51–60. Available from: `https://dl.acm.org/doi/10.1145/294469.294486`.

37. Nascimento MA, Sander J, Pound J. Analysis of SIGMOD's co-authorship graph. ACM SIGMOD Record. 2003;32(3):8–10. doi:10.1145/945721.945722.

38. Wang H, Hernandez JM, Van Mieghem P. Betweenness centrality in a weighted network. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics. 2008;77(4):046105.

39. Zhang P, Wang T, Yan J. PageRank centrality and algorithms for weighted, directed networks. Physica A: Statistical Mechanics and its Applications. 2022;586:126438.

40. Chen Z, Tang J, Wang G, Li S, Li X, Ji X, et al. UW-SDF: Exploiting Hybrid Geometric Priors for Neural SDF Reconstruction from Underwater Multi-view Monocular Images. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2024. p. 14248–14255. Available from: https://ieeexplore.ieee.org/document/10802499/?arnumber=10802499.

41. Shi J, Bai C, He H, Han L, Wang D, Zhao B, et al. Robust Quadrupedal Locomotion via Risk-Averse Policy Learning. In: 2024 IEEE International Conference on Robotics and Automation (ICRA); 2024. p. 11459–11466. Available from: https://ieeexplore.ieee.org/document/10610086/?arnumber=10610086.

42. Jia Y, Chen Y, Liu H, Li X, Li X. Hierarchical Learning and Control for In-Hand Micromanipulation Using Multiple Laser-Driven Micro-Tools. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2022. p. 1047–1054. Available from: https://ieeexplore.ieee.org/document/9982033/?arnumber=9982033.

43. Guo W, Meng C, Yuan E, He Z, Guo H, Zhang Y, et al. Compressed Interaction Graph based Framework for Multi-behavior Recommendation. In: Proceedings of the ACM Web Conference 2023. Austin TX USA: ACM; 2023. p. 960–970. Available from: https://dl.acm.org/doi/10.1145/3543507.3583312.

44. Zhu L, Zhang K, Chen H, Wei C, Zhang W, Tang H, et al. HCL4QC: Incorporating Hierarchical Category Structures Into Contrastive Learning for E-commerce Query Classification. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. Birmingham United Kingdom: ACM; 2023. p. 3647–3656. Available from: https://dl.acm.org/doi/10.1145/3583780.3614907.

45. Zhu X, Jin M, Zhang H, Meng C, Zhang D, Li X. Modeling Domains as Distributions with Uncertainty for Cross-Domain Recommendation. In: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. Washington DC USA: ACM; 2024. p. 2517–2521. Available from: https://dl.acm.org/doi/10.1145/3626772.3657930.

46. He C, Li K, Zhang Y, Tang L, Zhang Y, Guo Z, et al. Camouflaged Object Detection with Feature Decomposition and Edge Reconstruction. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE; 2023. p. 22046–22055. Available from: https://ieeexplore.ieee.org/document/10203727/.

47. Yang K, Tao J, Lyu J, Ge C, Chen J, Shen W, et al. Using Human Feedback to Fine-tune Diffusion Models without Any Reward Model. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE; 2024. p. 8941–8951. Available from: https://ieeexplore.ieee.org/document/10656897/.

48. Xiao Y, Luo Z, Liu Y, Ma Y, Bian H, Ji Y, et al. Bridging the Gap: A Unified Video Comprehension Framework for Moment Retrieval and Highlight Detection. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition

(CVPR). Seattle, WA, USA: IEEE; 2024. p. 18709–18719. Available from: https://ieeexplore.ieee.org/document/10656953/.

49. Pei S, Chari P, Wang X, Yang X, Kadambi A, Zhang Y. ForceSight: Non-Contact Force Sensing with Laser Speckle Imaging. In: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. Bend OR USA: ACM; 2022. p. 1–11. Available from: https://dl.acm.org/doi/10.1145/3526113.3545622.

50. Zhang H, Ba Y, Yang E, Mehra V, Gella B, Suzuki A, et al. WeatherStream: Light Transport Automation of Single Image Deweathering. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE; 2023. p. 13499–13509. Available from: https://ieeexplore.ieee.org/document/10205083/.

51. Zhou S, Chang H, Jiang S, Fan Z, Zhu Z, Xu D, et al. Feature 3DGS: Supercharging 3D Gaussian Splatting to Enable Distilled Feature Fields. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE; 2024. p. 21676–21685. Available from: https://ieeexplore.ieee.org/document/10656650/.

52. Singh AD, Ba Y, Sarker A, Zhang H, Kadambi A, Soatto S, et al. Depth Estimation from Camera Image and mmWave Radar Point Cloud. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE; 2023. p. 9275–9285. Available from: https://ieeexplore.ieee.org/document/10204897/.

53. Peters H, Ba Y, Kadambi A. pCON: Polarimetric Coordinate Networks for Neural Scene Representations. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE; 2023. p. 16579–16589. Available from: https://ieeexplore.ieee.org/document/10203580/.

54. Ezhov V, Park H, Zhang Z, Upadhyay R, Zhang H, Chandrappa CC, et al. All-day Depth Completion. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2024. p. 9425–9431. Available from: https://ieeexplore.ieee.org/document/10802376/?arnumber=10802376.

55. Yang Z, Wang J, Gan Z, Li L, Lin K, Wu C, et al. ReCo: Region-Controlled Text-to-Image Generation. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE; 2023. p. 14246–14255. Available from: https://ieeexplore.ieee.org/document/10203454/.

56. Aflalo E, Du M, Tseng SY, Liu Y, Wu C, Duan N, et al. VL-InterpreT: An Interactive Visualization Tool for Interpreting Vision-Language Transformers. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE; 2022. p. 21374–21383. Available from: https://ieeexplore.ieee.org/document/9880368/.

57. Li Z, Lu S, Guo D, Duan N, Jannu S, Jenks G, et al. Automating code review activities by large-scale pre-training. In: Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. Singapore Singapore: ACM; 2022. p. 1035–1047. Available from: https://dl.acm.org/doi/10.1145/3540250.3549081.

58. Lin Z, Gong Y, Liu X, Zhang H, Lin C, Dong A, et al. PROD: Progressive Distillation for Dense Retrieval. In: Proceedings of the ACM Web Conference 2023. Austin TX USA: ACM; 2023. p. 3299–3308. Available from: https://dl.acm.org/doi/10.1145/3543507.3583421.

59. Sun H, Liu X, Gong Y, Dong A, Lu J, Zhang Y, et al. LEAD: Liberal Feature-based Distillation for Dense Retrieval. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining. Merida Mexico: ACM; 2024. p. 655–664. Available from: https://dl.acm.org/doi/10.1145/3616855.3635774.

60. Ji Z, Ma P, Wang S. Perfce: Performance Debugging on Databases with Chaos Engineering-Enhanced Causality Analysis. In: 2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE); 2023. p. 1454–1466. Available from: https://ieeexplore.ieee.org/document/10298374/?arnumber=10298374.

61. Ji Z, Ma P, Wang S, Li Y. Causality-Aided Trade-Off Analysis for Machine Learning Fairness. In: 2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE); 2023. p. 371–383. Available from: https://ieeexplore.ieee.org/document/10298485/?arnumber=10298485.

62. Ma P, Ding R, Fu Q, Zhang J, Wang S, Han S, et al. Scalable Differentiable Causal Discovery in the Presence of Latent Confounders with Skeleton Posterior. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Barcelona Spain: ACM; 2024. p. 2141–2152. Available from: https://dl.acm.org/doi/10.1145/3637528.3672031.

63. Ma P, Ding R, Dai H, Jiang Y, Wang S, Han S, et al. ML4S: Learning Causal Skeleton from Vicinal Graphs. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Washington DC USA: ACM; 2022. p. 1213–1223. Available from: https://dl.acm.org/doi/10.1145/3534678.3539447.

64. Tian Y, Willis KDD, Al Omari B, Luo J, Ma P, Li Y, et al. ASAP: Automated Sequence Planning for Complex Robotic Assembly with Physical Feasibility. In: 2024 IEEE International Conference on Robotics and Automation (ICRA); 2024. p. 4380–4386. Available from: https://ieeexplore.ieee.org/document/10611595?denied=.

65. OpenAI. GPT-4 Turbo; 2024. Available from: https://openai.com/research/gpt-4-turbo.

66. Morrison PS, Dobbie G, McDonald FJ. Research collaboration among university scientists. Higher Education Research & Development. 2003;22(3):275–296.