Shaping Future Innovators: A Curriculum Comparison of Data Science Programs in Leading U.S. and Chinese Institutions

Dr. Elizabeth Milonas, New York City College of Technology

Elizabeth Milonas is an Associate Professor with the Department of Computer Systems at New York City College of Technology - City University of New York (CUNY). She currently teaches relational and non-relational databases and data science courses to undergraduate students. She holds a BA in Computer Science and English Literature from Fordham University, an MS in Information Systems from New York University, and a Ph.D. from Long Island University. Her research interests focus on three key areas: data science curriculum and ethics, retention of minority students in STEM degree programs, and organization and classification of big data.

Dr. Qiping Zhang, Long Island University

Dr. Qiping Zhang is an Associate Professor in the Palmer School of Library and Information Science at the C.W. Post Campus of Long Island University, where she also serves as director of the Usability Lab. Dr. Zhang holds a Ph.D. and an M.S. in information and library studies from the University of Michigan, Ann Arbor, and an M.S. and a B.S. in cognitive psychology from Peking University in Beijing, China. Prior to joining Long Island University in 2006, she worked at Drexel University, IBM Waterson Research Center, and Institute of Psychology at Chinese Academy of Science.

Dr. Zhang's general research areas are human-computer interaction (HCI), knowledge management (KM), social informatics and distance learning. Her primary interests lie in the areas of computer-supported cooperative work (CSCW) and computer-mediated communication. Specifically, she is interested in facilitating productive collaborations of individuals who are geographically and culturally distributed.

Dr. Duo Li, Shenyang Institute of Technology

Dr. Duo Li is an associate professor of Big Data Management and Application major at the School of Economics and Management, Shenyang Institute of Technology. Duo Li is a member of ASIST&T, and his research interests are focused on Human-Computer-Interaction, Big Data, and Data Analytics.

Shaping Future Innovators: A Curriculum Comparison of Data Science Programs in Leading U.S. and Chinese Institutions

1. Introduction

The Data Science field has been evolving rapidly both in the United States and in China in recent years. More and more day-to-day and business applications are depending on data science technologies such as data mining, machine learning, data management, and artificial intelligence [1]. With this rise in such data science technologies and tools, the demand for highly skilled data scientists has also grown exponentially [1]. According to Indeed Career Guide, data science related jobs were on the list of top 20 jobs in the United States in 2023[2]. These highly skilled professionals are responsible for complex tasks and have a pivotal role in organizations. Their effectiveness depends on technical skill, analytical proficiency and foundational understanding of all aspects related to the data science domain [1]. To meet the demand of training highly skilled and specialized Data Science professionals, many colleges have revised their existing majors to include Data Science related topics or created new Data Science related majors to focus on providing the Data Science knowledge and skills required for students to transform to effective Data Science professionals [3]. The goal of this research paper is to identify knowledge and skills in the core curriculum in Data Science undergraduate degree programs at leading colleges in the U.S. and China.

The research questions for this study are:
RQ1. *What are the knowledge and skills in the core curriculum of Data Science undergraduate degree programs in the U.S. and China?*
RQ2. *What are the thematic trends in the core curriculum of Data Science undergraduate degree programs in the U.S. and China?*

Understanding the similarities and differences between top data science curricula including key knowledge, skills and thematic trends in the U.S. and China is crucial for several reasons including curriculum design and best practices, alignment with industry needs, pedagogical approaches and learning outcomes, global competitiveness and workforce mobility, innovation and research impact, international collaboration and standardization and finally policy and educational reform.

In terms of curriculum design and best practices, comparing the structure, core courses, and specializations of data science programs in both countries can help identify best practices. This allows institutions to refine their curricula by incorporating successful approaches from each system, leading to more well-rounded and effective programs.

The alignment with industry need is crucial in both the U.S. and China and yet both countries have distinct yet overlapping demands for data science expertise. Examining each country's curricula can shed light on how each country prepares students for industry needs, ensuring that graduates possess relevant skills for today's global job market.

The pedagogical approaches and learning outcomes for each county are also unique. Differences in teaching methodologies, such as theoretical vs. applied focus, interdisciplinary integration, and use of emerging technologies, can offer valuable insights into how students are trained to solve real-world data problems. Understanding these approaches can help educators optimize teaching strategies.

As data science becomes increasingly global, students and professionals frequently transition between educational institutions and job markets across countries. A comparative analysis of global competitiveness and workforce mobility in each country can help ensure that graduates are equipped with competencies that allow them to thrive in diverse environments.

Both the U.S. and China contribute significantly to data science research. Examining how their curricula emphasize research, innovation, and ethics can reveal how each system nurtures future thought leaders and technological advancements in the field.

Understanding curriculum differences between the U.S. and China can facilitate student exchange programs, joint degree offerings, and research collaborations. It can also contribute to discussions on standardizing data science education to improve global interoperability.

Insights from comparison of international collaboration and standardization in both countries can guide policymakers and academic leaders in shaping future data science education policies, ensuring that programs remain relevant in a rapidly evolving technological landscape.

By studying the similarities and differences in both countries' policy and education reform, institutions can enhance their curricula, better prepare students for global opportunities, and contribute to the advancement of data science education on an international scale.

2. Literature Review

The Data Science field is multi-faceted and interdisciplinary [4,5] encompassing not only computer programming and statistics but also social and management aspects [4]. Both the United States and China have been forerunners in the data science domain developing and enhancing data science related tools and technologies [3,4]. Higher education in both countries has also been expanding and growing alongside the Data Science discipline [6]. In a recent study [7], an analysis of data science curriculum content and program characteristics offered by various ranking colleges in China and the United States showed that converging trends are evident despite the divergent national strategies employed by these two countries. These study findings [7] highlight the significance of data science education in both countries and show their shared recognition of its importance.

The knowledge and skill set data scientists need to succeed as proficient professionals in this dynamic domain is multi-dimensional [8]. A recent study [8] outlined the various knowledge and skills required for a proficient data scientist which include technical proficiency, analytical thinking and problem solving, domain-specific knowledge, and continuous learning and adaptability [8]. The results of a content analysis of 30 randomly selected Data Science programs [9] showed that most Data Science courses covered the basic level of analytical skills required, however, upper-level skills were not efficiently covered. In a more recent study [10] where 60

data science courses syllabi across all universities in Taiwan was examined, the results show a focus on moderate complexity learning objectives. The study results [10] showed the topics covered focused on applying and analyzing cognitive processes and included four key topics: Big Data, Analytical techniques, programming, and teaching strategies.

2.1 *Comparison of University Ranking Agencies in the U.S.*

A review of college ranking agencies was conducted to determine the most suitable source for use in this study. Four prominent agencies were identified, and relevant details for each are presented in Table 1 below. The selected ranking agencies include:

- U.S. News & World Report is often regarded as the benchmark for college rankings, with a strong emphasis on institutional prestige and academic reputation.
- Forbes prioritizes financial outcomes and post-graduate success, aligning well with analyses focused on return on investment.
- Niche offers insights into student experiences, campus life, and overall satisfaction, making it a valuable resource for assessing student-centered outcomes.
- Washington Monthly ranks institutions based on their contributions to public service, social mobility, and affordability, thus providing a broader perspective on institutional impact.

Established in 1983, U.S. News & World Report utilizes a multifaceted methodology that includes graduation and retention rates, faculty resources, student selectivity, expert opinion, and financial resources. Its long-standing presence and comprehensive approach have led it to be considered the standard in college ranking systems.

Table 1: Comparison of College Ranking Agencies

Ranking	Primary Focus	Key Metrics	Best for Students Prioritizing	Strengths	Weaknesses
US News Ranking	Academic Reputation, Selectivity, Faculty Quality	Graduation rates, faculty credentials, peer reputation, financial resources, student selectivity	Prestige, faculty excellence, and school reputation	Highly respected and widely used in academia	Favors wealthier schools, does not consider student experience much
Forbes Ranking	ROI (Return on Investment), Career Success	Alumni salaries, student debt, graduation rates, leadership outcomes, student satisfaction	High financial return, career prospects, and earnings	Strong focus on financial outcomes and long-term success	Overlooks campus life, student diversity, and faculty resources
Niche Ranking	Student Experience, Campus Life, Social Aspects	Student and alumni reviews, academics, diversity, athletics, campus quality	A well-rounded college experience, including social life and student satisfaction	Includes real student feedback and non-academic factors	Less emphasis on financial outcomes and long-term career success
Washington Monthly	Social Mobility, Public Service, Research Impact	Graduation rates of Pell Grant recipients, research funding, community service participation	Affordability, societal impact, and upward mobility	Highlights colleges that support low-income	Less emphasis on prestige and post-graduation earnings

				students and public service	

2.2 Top-Ranking Chinese and U.S. Colleges

An earlier study [11] identified the Data Science undergraduate programs in the U.S and China based on college ranking as shown in Table I and II.

TABLE I
U.S. NEWS RANKING CATEGORIES, NUMBER OF COLLEGES AND THE RANK RANGES FOR EACH CATEGORY

U.S. News Ranking Category	N of colleges	Ranking Ranges
National	66	1: 1-75 2: 76-150 3: 151-225 4: 226-297 5: 298-389
Regional	36	1: 1-16 2: 17-75 3: 76-150
National Liberal Arts	8	1: 15 2: 84 3: 102 4: 113, 5: 113 6: 120 7: 127 8: 140

TABLE II
SHANGHAI CONSULTANCY RANKING OF DATA SCIENCE PROGRAMS IN CHINA

Ranking	2021	2022	2023
A+	14	17	18
A	59	72	79
B+	65	91	189
B	211	263	194
Total	**309**	**352**	**370**

2.2 ACM Data Science Competencies

Results from the earlier study [7,12] identified four core ACM Data Science competencies as the most prevalent in top-tier Chinese and U.S. undergraduate data science programs ((N). These top four ACM competencies are: Computing Fundamentals, Data Management/Governance/Privacy), Data Mining and Big Data, and Machine Learning.

TABLE III
U.S. DATA SCIENCE PROGRAM NATIONAL RANKING (NR) AND PROGRAM COMPETENCIES

Competencies	U.S News National Ranking

	NR1	NR2	NR3	NR4	NR5	Total N
Machine Learning	**50%**	24%	19%	2%	5%	42
Computing Fundamentals	**38%**	29%	18%	6%	8%	61
Data Management/ Governance/ Privacy	**38%**	28%	16%	8%	10%	50
Data Mining/Big Data	**38%**	25%	19%	6%	11%	47
Math & Stats	37%	31%	17%	8%	8%	65
Data Science in Context	36%	32%	18%	7%	7%	44
Data Visualization	33%	30%	22%	8%	5%	36
Total	133	99	63	23	27	345

2.3 Data Science Knowledge and Skills Framework

Table IV shows a knowledge and skills framework for evaluating Data Science proficiency. Vinay [8] categorized technical, analytical, and domain-specific knowledge and skills into four categories: Knowledge domain, Technical Skills, Analytical Skills and Domain-specific knowledge. The table below identifies the details for each of these categories.

TABLE IV
DATA SCIENCE KNOWLEDGE AND SKILL FRAMEWORK

Main Category	Sub-category
1.0 Knowledge Domain	1.1 Data Structures and algorithms 1.2 Basic statistics 1.3 Database management 1.4 Data science principles 1.5 Machine learning models 1.6 Big data frameworks
2.0 Technical Skills	2.1 Programming languages: Python, R, SQL 2.2 Data visualization tools: Tableau, Power BI, Matplotlib, seaborn 2.3 Data Processing Frameworks: Apache Hadoop, Apache Spark 2.4 Machine learning libraries: TensorFlow, Scikit-learn, PyTorch 2.5 Database Management Systems: MySQL, PostgreSQL, MongoDB 2.6 Version Control Systems: Git 2.7 Big Data Technologies: Apache Kafka, Apache Kafka 2.8 Cloud Platforms: AWS (Amazon Web Services), Microsoft Azure, Google Cloud Platform 2.9 Integrated Development Environments: Jupyter Notebooks, RStudio
3.0 Analytic Skills	3.1 Data Cleaning and Preprocessing 3.2 Feature Engineering and Model Evaluation

	3.3 Problem Formulation 3.4 Hypothesis Generation 3.5 Data Exploration 3.6 Statistical Analysis 3.7 Machine Learning Application 3.8 Iterative Refinement 3.9 Critical Thinking 3.10 Optimization Strategies 3.11 Interdisciplinary Collaboration 3.12 Continuous Learning
4.0 Domain-specific Knowledge	4.1 Industry Contextualization 4.2 Relevant Data Variables 4.3 Customized Modeling Approaches 4.4 Understanding Business Objectives 4.5 Data Privacy and Compliance 4.6 Effective Communication with Stakeholders 4.7 Identification of Key Performance Indicators 4.8 Adaptability to Industry Trends 4.9 Problem Solving Relevance 4.10 Strategic Decision Support

3. Method:

Program Choice: U.S. News & World Report was chosen because it has published college rankings since 1983 and is often considered the industry standard. Its methodology incorporates a range of indicators including graduation and retention rates, faculty resources, student selectivity, expert assessments, and financial resources.

This study focused on the top-tier Chinese and U.S. colleges identified in [7]. Chinese colleges ranked A or A+ by the Shanghai Consultancy Ranking and U.S. colleges ranked as category 1 by U.S. News National Ranking, were chosen for the content analysis. However not all programs list their course syllabus online. Only those who provide free public access to the course syllabus were chosen for the study.

Table V lists all 14 data science programs chosen for this study. Six U.S. institutions analyzed including three private institutions: Columbia University, Duke University, University of Rochester, and three public institutions: University of Georgia, University of San Diego, and University of Wisconsin-Madison. Eight Chinese institutions analyzed were Beijing Institute of Technology, Central University of Finance and Economics, Harbin Institute of Technology, Hefei University of Technology, Peking University, Shandong University of Finance and Economics, Southwestern University of Finance and Economics, and University of Electronic Science and Technology of China.

China	USA
1. Beijing Institute of Technology 2. Central University of Finance and Economics 3. Harbin Institute of Technology 4. Hefei University of Technology 5. Peking University 6. Shandong University of Finance and Economics 7. Southwestern University of Finance and Economics 8. University of Electronic Science and Technology of China	1. Columbia University 2. Duke University 3. University of Georgia 4. University of Rochester 5. University of San Diego 6. University of Wisconsin-Madison

For each selected data science program, both the program affiliation and the associated school or college were recorded. Table VI presents a comparison of these affiliations between Chinese and American programs. Programs in both countries commonly fall under the domains of computer science and data science. However, interesting differences exist: Chinese programs more frequently include Big Data Management, whereas American programs uniquely feature interdisciplinary studies and those based in Mathematics and Statistics. Regarding school or college affiliation, further distinctions emerge. In the U.S., 50% of programs are housed within Colleges of Arts and Sciences, compared to 0% in China. Conversely, in China, 25% of programs are affiliated with Schools of Computer Science and 62.5% with Schools of Management—categories not represented among the American programs.

TABLE VI
NUMBER OF CORE CURRICULUM COURSES BY COUNTRY

	China	USA
Program Affiliation		
Big Data Management	5	0
Computer Science	1	2
Data Science	2	2
Interdisciplinary	0	1
Math & Statistics	0	1
Total: Program	**8**	**6**
School/College		
Arts & Sciences	0	3
College of Engineering	1	1
Data Science Institute	0	1
School of Computer Science	2	0
School of Management	5	0
N/A	0	1
Total: School/College	**8**	**6**

Curriculum Choice: For all chosen data science programs, we chose the syllabi from core data science curriculum for our content analysis. Core curriculum is determined by whether it covers the top competencies identified by previous study [18] and listed in above Table III.

Tables VII and VIII present the number of core data science courses selected for this study, organized by country and by institution, respectively. Both tables demonstrate that the Chinese and U.S. program samples are comparable in terms of competency coverage and the balance between required and elective courses. Table VIII further highlights variations between institutions, which may reflect either broader curricular options or differences in syllabus availability. In this study, "core curriculum" refers specifically to data science-related courses offered as part of the degree program, excluding general education requirements. To support the representativeness of our sample, we report the total number of required and elective courses for each program, alongside the number of courses included in our analysis for each category.

TABLE VII
NUMBER OF CORE CURRICULUM COURSES BY COMPETENCIES AND COUNTRIES

Top Competencies	China	USA
Computing Fundamentals	18	18
Data Management	13	10
Data Mining	4	3
Machine Learning	7	9
Total	**42**	**40**

TABLE VIII
NUMBER OF SELECTED CORE CURRICULUM COURSES BY UNIVERSITY

University	Required Courses: selected (total)	Elective Courses: selected (total)	Total Selected Courses
Beijing Institute of Technology	3 (18)	4 (9)	7
Central University of Finance and Economics	0 (10)	6 (23)	6
Harbin Institute of Technology	0 (7)	3 (18)	3
Hefei University of Technology	4 (18)	0 (25)	4
Peking University	3 (9)	3 (15)	6
Shandong University of Finance and Economics	1 (6)	4 (15)	5
Southwestern University of Finance and Economics	0 (14)	5 (29)	5
University of Electronic Science and Technology of China	6 (7)	n/a	6
Total - China	**17**	**25**	**42**
Columbia	0 (13)	5 (5)	5
Duke	8 (9)	0 (3)	8
University of Rochester	3 (8)	0 (3)	3
University of Georgia	4 (11)	4 (4)	8
University of San Diego	6 (8)	6 (6)	12
University of Wisconsin-Madison	2 (8)	2 (6)	4
Total - U.S.	**23**	**17**	**40**

Table VIX listed the course titles by top four competencies in both China and U.S.A..

China	USA
Computing Fundamentals	
1. Advanced Programming Language (Python) 2. Algorithm Analysis and Design 3. Big Data Algorithms 4. Big Data Analysis Principles and Applications 5. Big Data Fundamentals and Applications 6. Computer Networks 7. Computer Operating Systems 8. Computer Systems and Network Security 9. Data Structure and Algorithm(A) 10. Data Structures and Algorithms 11. Digital Economics 12. Introduction to Big Data Technologies 13. Introduction to Computer Systems 14. Introduction to E-commerce* 15. Principles and Practice of Parallel Programming 16. Python and Financial Big Data Analysis Fundamentals 17. Python Programming	1. Algorithm and Data Structure COMPSCI 201 2. Computer Science Theory COMS W3261 3. Data Science Programming I CS 220 4. Data Science Programming II CS 320 5. Data structures and algorithms DSC 30 6. Design and Analysis COMPSCI 330 7. Foundations for Informatics & Data Analytics CSCI 1360 8. Introduction to Computational Complexity COMS 4236 9. Introduction to Computational Learning Theory COMS 4252 10. Introduction to Computational Science CSCI 2150 11. Introduction to Computer Systems CPS 210 12. Introduction to Parallel Programming COMS 4130 13. Principles of Data Science DSC10 14. Programming and Data Structures DSC 20 15. Software Development CSCI 1203 16. Systems for Scalable Analytics DSC102 17. Theoretical foundations of DS DSC 40A 18. Theoretical Foundations of DS II DSC 40B
Data Management	
1. Big Data and Finance 2. Big Data Computing Technologies 3. Big Data Marketing 4. Big Data Thinking and Technology 5. Database Principles and Applications 6. Digital Marketing 7. Financial Technology 8. Information Economics 9. Introduction to Database Systems 10. Management Information Science 11. Management Information Systems* 12. Marketing	1. Data Science COGS9 2. Data Science I CSCI 3360 3. Data Structures CSCI 2720 4. Database Management CSCI 4370 5. Database Systems CSC 261/461 6. Everything Data CS 216 7. Introduction to Data Management DSC 100 8. Introduction to Database * COMPSCI 316/COMS4111 9. Practice and Applications of DS DSC 80
Data Mining	
1. Data Mining and Python Practice 2. Data Warehouse and Data Mining 3. Financial Data Mining (Bilingual) 4. R Language and Financial Data Mining	1. Data Mining CSC/DSC 240/440 2. Data Mining CSCI 4380 3. Introduction to Data Mining DSC 148
Machine Learning	
1. Artificial Intelligence and Computational Science 2. Artificial Intelligence: Deep Learning 3. Business Intelligence 4. Fundamentals of Deep Learning 5. Introduction to Artificial Intelligence 6. Machine Learning*	1. Artificial Intelligence * COMPSCI 570 & CS 3540 2. Data Science II CSCI 4360 3. Elements of Machine Learning COMPSCI 371 4. Introduction to Artificial Intelligence COMPSCI 370 5. Machine Learning DSCC/CSC 242 6. Matrix Methods in Machine Learning CS 532

	7. Probabilistic modeling and machine learning DSC140A
	8. Representation Learning DSC1 40B

Content Analysis: This study utilized content analysis to examine the core Data Science course syllabi from top-ranking colleges in China and the United States. The analysis focused on identifying the knowledge and skills outlined in [7] as well as the ACM Data Science competencies described in [15,16,18]. By systematically reviewing course descriptions, topics, assignments, and materials, the study aimed to uncover trends and differences in how foundational and advanced Data Science concepts are integrated into the curricula of these leading institutions.

A total of 40 U.S. course syllabi were examined, and content analysis was conducted using course titles, course description, weekly topics, homework assignments, project descriptions, and classwork. Similarly, a total of 42 Chinese course syllabi were examined and content analysis was conducted using course title, course description, and course topics.

Three domain experts developed codes based on ACM curricular framework. They coded the sample syllabus data, achieving an acceptable inter-coder reliability with over 85% agreement.

4. Results

4.1 Comparison of Data Science Knowledge and Skill in Core Curriculum

Based on the framework for data science knowledge/skill in Table IV, we coded course topics for all 82 courses. Table IX lists the total number of course topics for each of core competencies.

Table IX
NUMBER OF COURSE TOPICS BY COMPETENCIES

Competencies	China	USA
Computing Fundamentals	169	366
Data Management	139	193
Data mining	43	54
Machine learning	77	229
Total	**428**	**842**

The data science curriculum encompasses a total of 37 knowledge and skill categories [8]. Table X highlights only the top three categories within each of the four core competencies selected for analysis in this study, which were identified based on prior research [7, 12]. These results are derived from an analysis of course topics rather than course titles, allowing for a more detailed and accurate representation of the knowledge and skills emphasized within the curricula.

For the competency of computing fundamentals, knowledge of *1.1 Data Structures and Algorithms* ranked as the top priority in both China and the U.S. Both countries also emphasized the technical skill of *2.1 Programming Languages*. Additionally, China prioritized knowledge of 1.3 Database Management, while the U.S. focused on 1.4 Data Science Principles.

For the competency of data management, knowledge of *1.3 Database management* and *1.4 Data science principles* ranked highly in both Chinese and U.S. core curricula. However, Chinese courses placed greater emphasis on knowledge of *1.1 Data Structures and Algorithms*, while U.S. courses prioritized the skill of *2.1 Programming Languages* within this competency.

For the competency of data mining, while Chinese programs predominantly focused on the knowledge of *1.1 Data Structures and Algorithms* (72%), In contrast, U.S. programs maintained the same priorities as in the competency of data management: *1.3 Database management, 1.4 Data science principles, and 2.1 Programming language*.

For the competency of machine learning, the knowledge of *1.1 Data structure and algorithms* remained as the predominate focus in Chinese programs, ranking as the top priority (57%). In both Chinese and U.S. programs, knowledge of *1.5 Machine learning models* ranked highly among course topics. Additionally, the application of machine learning was emphasized, with Chinese programs ranked *3.3 Problem formulation* as the second priority, while U.S. programs ranked *3.7 machine learning application* as their second focus.

Interestingly, domain-specific knowledge did not rank as a top priority in either Chinese or U.S. programs. Additionally, the analytic skill of *3.6 statistical analysis* ranked as the third priority for the competencies of data mining and machine learning for Chinese programs only.

Table X
TOP THREE KNOWLEDGE/SKILLS FROM COURSE TOPICS BY COMPETENCIES IN CHINA AND USA

China			USA		
Computing Fundamentals					
1.1 Data structures and algorithms	91	54%	1.1 Data structures and algorithms	112	31%
1.3 Database management	17	10%	2.1 Programming languages…	63	17%
2.1 Programming languages…	11	7%	1.4 Data science principles	39	11%
Data Management					
1.1 Data structures and algorithms	49	35%	1.3 Database management	61	32%
1.3 Database management	15	11%	1.4 Data science principles	19	10%
1.4 Data science principles	13	9%	2.1 Programming languages…	17	9%
Data Mining					
1.1 Data structures and algorithms	31	72%	1.3 Database management	61	32%
2.4 Machine learning libraries…	6	8%	1.4 Data science principles	19	10%
3.6 statistical analysis	4	5%	2.1 Programming languages…	17	9%
Machine Learning					
1.1 Data structures and algorithms	44	57%	1.5 Machine learning models	19	35%
1.5 Machine learning models	13	17%	3.7 Machine learning application	14	26%
3.3 Problem formulation	3	4%	1.4 Data science principles	7	13%
3.6 statistical analysis	3	4%			

4.2 Analysis of Thematic Trends

First, we used open coding techniques to conduct content analysis on course descriptions to identify major themes in chosen core courses. Table XI summarized the major themes from both Chinese and U.S. course descriptions.

COMPARISON OF MAJOR THEMES IN CORE CURRICULUM BETWEEN CHINA AND USA

Chinese Curriculum	U.S. Curriculum
● Machine Learning and AI ● Big Data and Data Science ● Ethics and Privacy ● Programming and Algorithms ● Practical Skills Development ● Integration with Industry	● Machine Learning and AI ● Core Computational Competencies: ● Data Management and Mining ● Ethical Concerns ● Interdisciplinary Applications ● Theory and Practical Balance

Our analysis showed three common themes in Chinese and U.S. core data science courses. First, both curricula prioritize *Machine Learning and AI* as critical theme, emphasizing foundational models like neural networks and deep learning. Second, *Data Management and Mining* are fundamental, with courses on database systems, SQL, and Big Data analysis techniques. Third, ethical considerations surrounding *Data Privacy and AI Applications* are integral to both programs.

Our analysis also identified four differences among themes in Chinese and U.S. core data science courses (Table XII).

Table XII
DIFFERENCES IN CORE CURRICULUM BETWEEN CHINA AND USA

Aspect	Chinese Curriculum	U.S. Curriculum
Practicality	Industry-specific applications (fintech, marketing).	Broader applicability across diverse fields.
Programming Language	Python and C; less emphasis on object-oriented design.	Python and Java; strong focus on object-oriented methods.
Interdisciplinarity	Integration with finance, marketing, and accounting.	Emphasis on fields like healthcare, humanities, robotics.
Big Data Techniques	Application-driven; solving real-world industry challenges.	Scalability and distributed systems (e.g., MapReduce).

First is on *Practical Focus*. The Chinese curriculum places stronger emphasis on industry applications (e.g., fintech, marketing, and interdisciplinary domains), while the U.S. curriculum integrates these aspects but leans more towards theoretical foundations and broad applicability across disciplines.

Second is on *Programming Languages*. Both programs include Python, but the U.S. program emphasizes Java for object-oriented programming, while the Chinese curriculum uses C alongside Python for foundational skills.

Third is on *Interdisciplinary Approach*. The Chinese curriculum weaves business-oriented themes like financial technology and marketing deeply into coursework, whereas the U.S. curriculum incorporates broader applications like healthcare, humanities, and robotics.

Four is on *Big Data and Scalability*. The U.S. curriculum includes more emphasis on scalability and distributed systems (e.g., MapReduce), whereas the Chinese curriculum focuses more on applied techniques in financial and social contexts.

4.3 Implication of the Study
This study raises several important questions regarding the potential implications of the findings:
- How do different educational approaches affect students' preparedness for future roles in data science?
- Do these curricular differences reflect distinct educational goals in China and the U.S.? A more comprehensive analysis, including additional syllabi, would help explore this question further.
- In what ways might these approaches influence graduates' readiness for diverse career paths, such as positions in industry versus academia?
- The U.S. curriculum places greater emphasis on Java and object-oriented programming, while Chinese programs tend to use C along with Python. Although these differences may not significantly impact students' adaptability, they could shape programming proficiency in specific contexts.
- Could variations in interdisciplinary integration and the range of data science techniques taught influence graduates' capacity for cross-disciplinary collaboration?

These questions do not demand definitive answers within the current study. Instead, they serve to highlight the broader theoretical and practical relevance of the findings. Raising such questions can help guide future research and underscore the significance of this study's contributions.

4.4 Limitation of the Study
A key limitation of this study is that only institutions offering publicly accessible course syllabi were included in the analysis. This restriction may limit the generalizability of the findings. Future studies would benefit from direct outreach to institutions to obtain a more comprehensive set of syllabi, enabling a deeper and more representative analysis.

5. Conclusion

An analysis of undergraduate data science programs in the U.S. and China reveals that both curricula effectively equip students with essential competencies in mathematics and statistics, data management, data science technologies, and programming. Despite these shared core areas, the two countries demonstrate distinct educational priorities. The Chinese curriculum adopts a strongly application-driven approach, aligning closely with immediate industry needs. In contrast, U.S. programs emphasize theoretical depth, scalability, and broader interdisciplinary applications, preparing students for a wider range of professional and research pathways.
The study's identification of key knowledge areas and specific subtopics covered in each country's curriculum offers valuable insight into the structure and focus of data science education. In addition, thematic comparisons across the core curricula highlight differing emphases in application domains: Chinese programs tend to concentrate on finance, marketing, and accounting, while U.S. programs incorporate a broader spectrum of fields, reflecting a more diverse and research-oriented educational model.

6. Reference

[1]   Heidarysafa, M., Kowsari, K., Bashiri, M., & Brown, D. E. (2022). Towards a Knowledge Discovery Framework for Data Science Job Market in the United States. In *Proceedings of the Future Technologies Conference (FTC) 2021, Volume 1*(pp. 875-887). Springer International Publishing.

[2]   Gafner, J. (2023) Best Jobs of 2023. Indeed Career Guide, 12/19/2024. Accessed: 1/12/2024 https://www.indeed.com/career-advice/news/best-jobs-of-2023

[3]   Zhang, Y., Wu, D., Hagen, L., Song, I. Y., Mostafa, J., Oh, S., and Saltz, J. S. (2023). "Data science curriculum in the iField." Journal of the Association for Information Science and Technology, 74(6), 641- 662.

[4]   Cao, L. (2017). "Data science: a comprehensive overview." ACM Computing Surveys (CSUR), 50(3), 1-42.

[5]   Raban, D. R., & Gordon, A. (2020). "The evolution of data science and big data research: A bibliometric analysis." Scientometrics, 122(3), 1563-1581.

[6]   Miah, S. J., Solomonides, I., & Gammack, J. G. (2020). "A design-based research approach for developing data-focussed business curricula." Education and Information Technologies, 25(1), 553-581.

[7]   D. Li, E. Milonas and Q. Zhang. "Converging Paths in Divergent Systems: A Comparative Analysis of Data Science Education Strategies in China and the United States." 2024 Frontiers in Education(FIE), IEEE, Washington, DC. October 13-16, 2024.

[8]   Vinay, B. (2024). "Data Scientist Competencies and Skill Assessment: A Comprehensive Framework." International Journal of Data Scientist (IJDST) Volume 1, Issue 1, January-June 2024, pp. 1-11. Available: https://iaeme.com/MasterAdmin/Journal_uploads/IJDST/VOLUME_1_ISSUE_1/IJDST_01_01_001.pdf

[9]   Tang, Rong & Lim, Watinee. (2016). "Data science programs in U.S. higher education: An exploratory content analysis of program description, curriculum structure, and course focus." Education for Information. 32. 269-290. 10.3233/EFI-160977.

[10]   Hsu, Yu-Chia. (2024). "Mapping the Landscape of Data Science Education in Higher General Education in Taiwan: A Comprehensive Syllabi Analysis." Education Sciences 14(7), 763.

[11]   E. Milonas, Q. Zhang, and D. Li. "Do Undergraduate Data Science Program Competencies Vary by College Rankings?". 2022 ASEE Annual Conference & Exposition, Minneapolis, MN. Available: https://peer.asee.org/41153

[12]   E. Milonas, D. Li, and Q. Zhang, "Content analysis of two-year and four-year data science programs in the United States," presented at the 2021 ASEE Virtual Annual Conference, Virtual Conference, July 2021. [Online]. Available: https://peer.asee.org/36842.