IIAI Open Conference Publication Series
IIAI Letters on Institutional Research
Vol.004, LIR321
DOI: https://doi.org/10.52731/lir.v004.321

Analysis of Discrepancies in Learning Awareness of Data Science Across Disciplines - A Case Study of Nihon University, College of Humanities and Sciences

Eriko Tanaka *, Takaaki Ohkawauchi *

Abstract

In recent years, there has been a growing demand for data science education alongside rapid advancements in technology, particularly in AI. However, despite various attempts to innovate curricula and teaching methods at universities, there is limited research on students' awareness of their learning needs in this field. Hence, this study investigated the recognition of the importance of data science and willingness to take related courses among a diverse range of students, including those from both scientific and non-scientific disciplines. The results revealed that although 90% of the students recognized the necessity of data science in the future, less than half were considering taking related courses. In our faculty, we are considering measures, such as compulsory courses and certification systems, to enhance students' knowledge and skills in data science and make them more accessible to a wider range of students. This paper summarizes the background and findings of the study.

Keywords: data science, artificial intelligence, mathematics, curriculum

1 Introduction

The Japanese Cabinet Office has embraced the term 'Society 5.0' to describe "a human-centered society that harmonizes economic progress with addressing social challenges through a system that deeply integrates cyberspace and physical space [1]." This can be paraphrased as "a society where necessary goods and services are provided to the right people at the right time, in the right amount, able to finely address various societal needs, ensuring everyone receives high-quality services, transcending differences such as age, gender, region, language, enabling vibrant and comfortable living [2]." There is a need to transform the current society into an "AI-ready" society, which integrates AI technology to consider aspects such as individuals, social systems, industrial structures, innovation systems, and governance [3]. To realize such a society, it is no longer sufficient for only engineers or researchers to have expertise in AI, as has traditionally been the case. Instead, every member of society must become AI-ready.

One specific goal for all university students, regardless of their academic background, is to achieve literacy-level understanding or higher in mathematics, data science, and AI [4]. With this perspective in mind, the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) initiated the "Approved Program for Mathematics, Data Science and AI Smart Higher Education, designated by Gov of Japan (MDASH)" from the academic year 2021 for literacy

* Nihon University, Tokyo, Japan

level, and from the following academic year 2022 for applied literacy-plus level [5]. This program aims to expand nationwide, providing literacy-level education to 500,000 university and technical college students annually by the fiscal year 2025, along with literacy-plus-level education to 250,000 high school students. In the academic year 2021, there were approximately 630,000 university entrants [6], and technical colleges enrolled approximately 60,000 students [7]. Considering the short timeframe until the fiscal year 2025, it is evident that the country aims for a high proportion of AI-ready citizens nationwide (Figure 1).

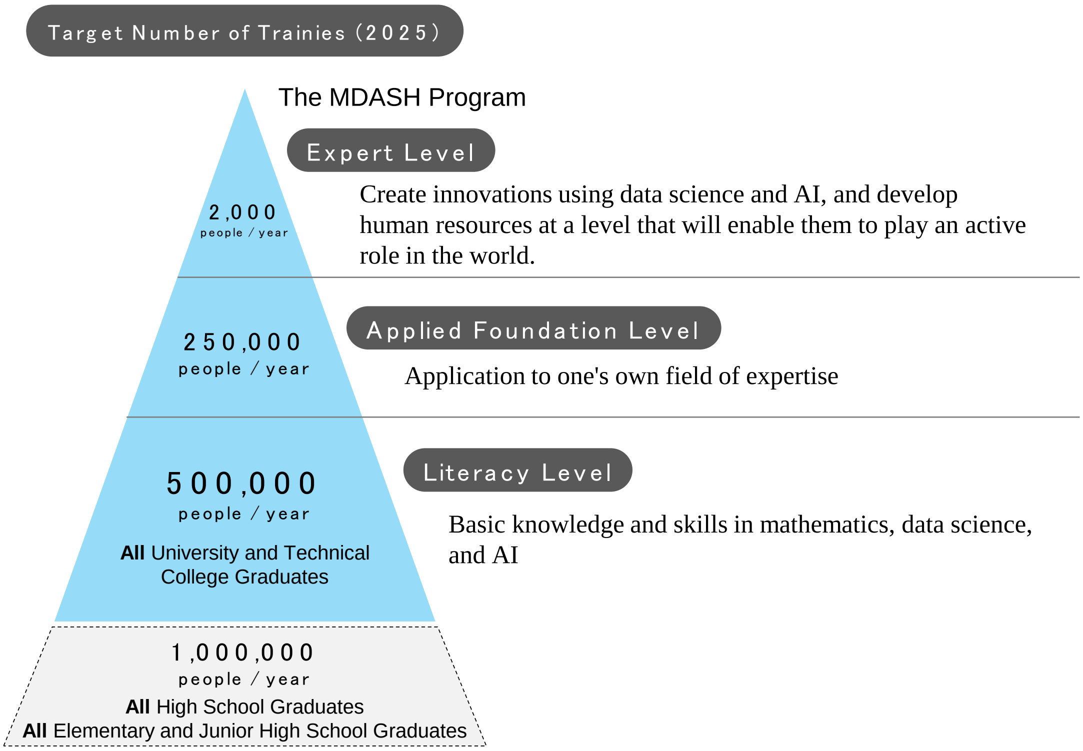

Figure 1: Approved Program for Data Science Education in Japan

In recent years, many universities have applied for accreditation to MDASH, and as of now, approximately 37.5%, or 297 out of a total of 793 universities, have already received accreditation at the literacy level. Not being accredited by the MDASH may be regarded as lagging in providing data science education at universities. Therefore, many other universities aim for accreditation to prevent this situation.

While the goal is to provide such education to all students, it is essential to recognize that the attitude of students, who are learners, is equally crucial, regardless of how enthusiastic the educators may be. If the significance and necessity of learning data science are not shared with learners, this could potentially impact the learning process. Hence, the purpose of this study was to investigate the extent to which students in various fields are aware of and perceive the importance of learning AI and data science.

2 Related Works

2.1 Definition of Data Science

When comparing the attention given to data science in Japan and the rest of the world, similar trends are observed. Figure 2 represents the indexing of the search terms "Data Science" and "Data Science (in Japanese)" on Google Trends [8] over 20 years from 2004 to 2023. The Y-axis represents the search frequency of each keyword at different times, with the highest monthly search volume for each keyword set at 100. Both were relatively low and stable until late 2012; however, since then, they have been consistently increasing. The end of 2012 is notable as the year when Davenport stated in the Harvard Business Review that "Data scientist is the sexiest job of the 21st century," which received media attention [9].

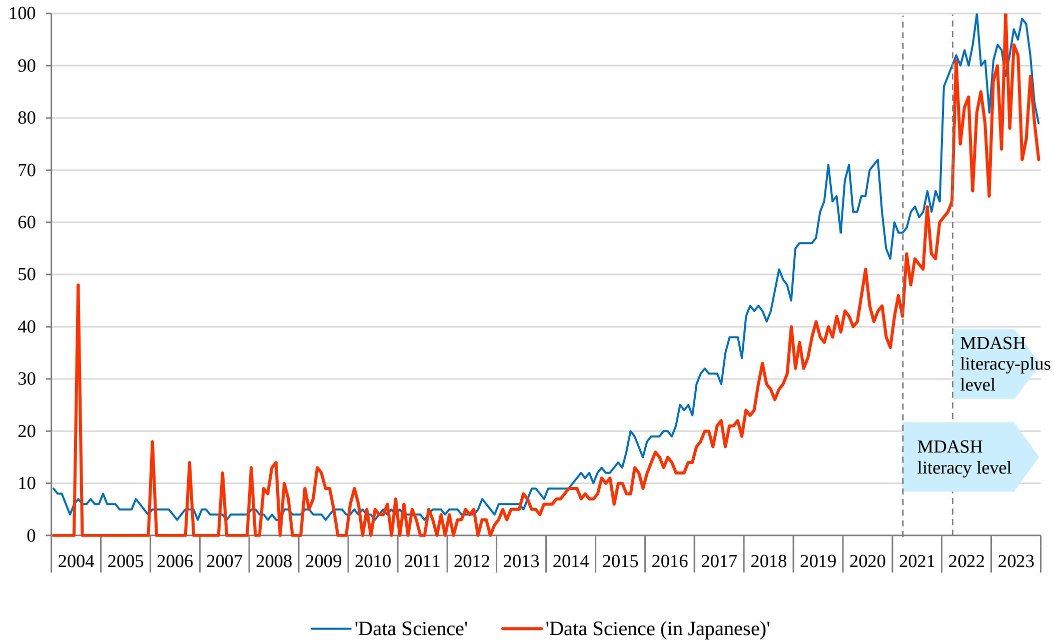

Figure 2: 'Data Science' Index by Google Trends

In Japan, data science trended slightly below the global trend until approximately 2021; however, since then, it has surged further and is closely aligned with the global trend. This can be attributed to the nationwide decision made in 2021 to promote data science education at the national level, as discussed in Section 1.

Before discussing data science education, it is important to clarify the definition of data science. One of the notable references where the term "Data Science" is clearly used is by Naur, a later Turing Award recipient, in 1974. While the definition of data science itself is not explicitly stated, it is emphasized that the principle of data science is the importance of concern regarding the characteristics of data processing tools. Furthermore, data science cannot be considered separately from statistics. In 1997, Jeff Wu, a chair of the Department of Statistics at the University of Michigan, proposed changing the name of "Statistics" as a discipline to "Data Science" [11].

In fact, the concept frequently utilized in recent definitions of data science, such as the one introduced in the first issue of the Harvard Data Science Review in 2019 as "the study of extracting value from data [12]," shares similarities with the first part of the American Statistical Association's definition of statistics: "Statistics is the science of learning from data and of measuring, controlling, and communicating uncertainty [13]."

2.2 Data Science Education

Data science education aims to cultivate data specialists and foster data literacy among citizens [14]. This definition aligns with Japan's MDASH, the data science education program certification system that aims to raise the data science capabilities of the entire population to a literacy level while nurturing a select group of top-level specialists. However, the interdisciplinary nature of data science education makes it challenging to define the boundaries of this field [15]. In addition to advocacy efforts in core disciplines, such as statistics, mathematics, computing, and artificial intelligence, which form the foundation of analytics, domain-specific data science is widely recognized [16]. Data science is increasingly permeating non-scientific domains such as law, history, nursing, media, and entertainment, with its relevance expanding beyond scientific fields [17][18].

What knowledge is required to acquire data science skills? According to several studies, it comprises three elements [19]: computing and data skills, math and statistics, and content. Data science is interdisciplinary in nature. Despite still underdeveloped aspects regarding its educational methods and content, the project "Data Science for Undergraduates" by The National Academies of Sciences, Engineering, and Medicine, an organization comprised of industry and university stakeholders in the United States, lists the following ten elements for undergraduate education in data science [20].

E1. Ethical problem-solving

E2. Data management and curation

E3. Data description and visualization

E4. Statistical foundations

E5. Mathematical foundations

E6. Computational foundations

E7. Data modeling and assessment

E8. Domain-specific considerations

E9. Communication and teamwork

E10. Workflow and reproducibility

STEM majors may have a certain level of understanding of this necessity. However, many non-STEM students might not only fail to comprehend the importance of skill sets but also struggle to envision the meaning of each skill set.

The current MDASH literacy level certifies data science education at universities, often based on a single model curriculum at each university. This means that regardless of their faculty or field of study, students learn from a single curriculum provided by their university. In this scenario, it is conceivable that there could be a mix of faculties where many students require mathematics and others where it is deemed unnecessary. Ideally, rather than teaching a uniform curriculum to diverse students, it would be preferable to vary the introductory methods and instructional designs by incorporating examples relevant to each student's field of study.

The aim of this study was to analyze the extent to which students' motivation for learning data science varies across their fields of study, as well as to examine the extent to which learning needs for the ten skill sets mentioned earlier differ.

3 Methods and Results of the Analysis

3.1 Curriculum of Our University

Nihon University has the largest number of university students in Japan, and its College of Humanities and Sciences (CHS) consists of three realms and 18 departments (Table 1), enrolling nearly 2,000 freshmen each year. The CHS has a scale equivalent to that of a comprehensive university and encompasses a wide range of specialized areas.

Table 1: College of Humanities and Sciences, Nihon University

Realms		Departments	
Humanities	2,744	D1: Philosophy	388
		D2: History	585
		D3: Japanese Language and Literature	598
		D4: Chinese Language and Culture	283
		D5: English Language and Literature	580
		D6: German Language and Literature	310
Social Sciences	3,539	D7: Sociology	921
		D8: Social Welfare	253
		D9: Education	522
		D10: Physical Education	940
		D11: Psychology	589
		D12: Geography	314
Natural Science	1,923	D13: Earth and Environmental Sciences	325
		D14: Mathematics	322
		D15: Information Science	342
		D16: Physics	259
		D17: Biosciences	298
		D18: Chemistry	377

* Number of students enrolled as of May 1, 2023

As for the subjects related to computer and data science shared among the faculties, until fiscal year 2019, there were two: Information Literacy and Basic Data Processing. Information Literacy is a foundational subject that covers various aspects of ICT, whereas Basic Data Processing is a subject in which students learn the fundamentals of statistics and methods for aggregation and visualization using Excel. Currently, only Information Literacy is a required course. In response to societal needs, a new course, “Big Data Science,” was established in 2020 to provide broad knowledge and skills related to AI. In addition, to ensure that all students within the CHS acquire literacy-level knowledge and skills in data science, Basic Data Processing and Big Data Science will become mandatory courses for enrollment (not required to earn credit for graduation) starting in 2025.

3.2 Learning Needs for Data Science

We first investigated the learning needs concerning the ten crucial elements outlined in Section 2 for data science education and examined the extent to which these needs differed among students and disciplines. The survey method involved distributing questionnaires during the initial class of the mandatory subject, Information Literacy, in academic year 2024. Students were requested to assess the necessity of the knowledge and skills associated with each item in the future, using a five-point scale ranging from "-2. Not necessary" to "2. Essential," or an option for "N/A. Unable to imagine what kind of knowledge or skills are involved."

The response rate for the survey was 1,799 of 2,126 (84.6%). Table 2 presents the results of the study. Each value in Table 2 represents the mean score on a five-point scale for the need for each data science learning element across the different academic departments. The mean values for all items in the table are above 0, indicating a general recognition of the need for learning elements of data science. The most significant difference between departments was evident in the awareness regarding "E1. Ethical problem-solving." Students in the science field or those with a higher interest in scientific knowledge and technology exhibited a better understanding of the necessity for ethical considerations. E2 and E3 pertain to knowledge of data-handling skills, with slightly higher overall scores observed among students in the science field. In contrast, students in the humanities, particularly those specializing in literature, tended to score lower. For "E4. Statistics" and "E5. Mathematics, " students in the humanities and social sciences understand their necessity in the future more than those in the natural sciences do. Furthermore, topping the list in all disciplines is "E8. Domain-specific considerations." There appears to be recognition across disciplines that incorporating data science into society requires knowledge and consideration of relevant domains. We argue that this figure reflects not only interest in data science education, but also a higher level of interest in their respective fields of specialization.

Table 2: Learning Needs for Each Element of Data Science

		E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
Departments	D1	0.42	1.03	1.14	1.76	1.68	1.45	1.51	1.76	1.61	1.18
	D2	0.72	1.38	1.45	1.73	1.78	1.67	1.70	1.88	1.77	1.26
	D3	0.51	1.16	1.35	1.74	1.78	1.68	1.57	1.79	1.74	1.20
	D4	0.50	1.05	0.78	1.59	1.84	1.49	1.54	1.74	1.54	1.25
	D5	0.77	1.31	0.98	1.55	1.90	1.71	1.64	1.78	1.77	1.41
	D6	0.45	1.25	1.03	1.58	1.78	1.54	1.52	1.71	1.57	1.06
	D7	0.83	1.50	1.30	1.66	1.76	1.77	1.74	1.85	1.72	1.38
	D8	0.77	1.42	1.11	1.69	1.87	1.56	1.80	1.89	1.73	1.12
	D9	1.06	1.41	1.28	1.58	1.80	1.76	1.75	1.78	1.72	1.41
	D10	0.80	1.22	1.06	1.39	1.73	1.61	1.47	1.55	1.55	1.29
	D11	0.94	1.56	1.41	1.76	1.79	1.62	1.61	1.90	1.73	1.18
	D12	0.72	1.38	1.33	1.62	1.69	1.62	1.68	1.76	1.68	1.17
	D13	1.23	1.54	1.44	1.47	1.62	1.65	1.72	1.86	1.74	1.26
	D14	1.62	1.54	1.44	1.49	1.68	1.79	1.67	1.75	1.68	1.55
	D15	1.42	1.48	1.31	1.57	1.60	1.90	1.64	1.86	1.76	1.44
	D16	1.42	1.47	1.44	1.59	1.86	1.80	1.61	1.73	1.65	1.51
	D17	1.28	1.56	1.43	1.56	1.72	1.61	1.67	1.75	1.61	1.40
	D18	1.48	1.53	1.38	1.23	1.54	1.66	1.65	1.71	1.60	1.38
Humanities		0.56	1.20	1.12	1.66	1.79	1.59	1.58	1.78	1.67	1.23
Social Sciences		0.85	1.42	1.25	1.62	1.77	1.66	1.67	1.79	1.69	1.26
Natural Science		1.41	1.52	1.41	1.48	1.67	1.74	1.66	1.78	1.67	1.42
Average		0.94	1.38	1.26	1.59	1.75	1.66	1.64	1.78	1.68	1.30

Furthermore, students were asked to rate on a five-point scale ranging from "2. Agree" to "-2. Disagree," whether they believe studying data science in university courses would be beneficial in their respective fields (Figure 3). There were slight variations depending on the field, but overall, positive opinions, including "2. Agree" and "1. Somewhat agree," totaled approximately 90% in all disciplines.

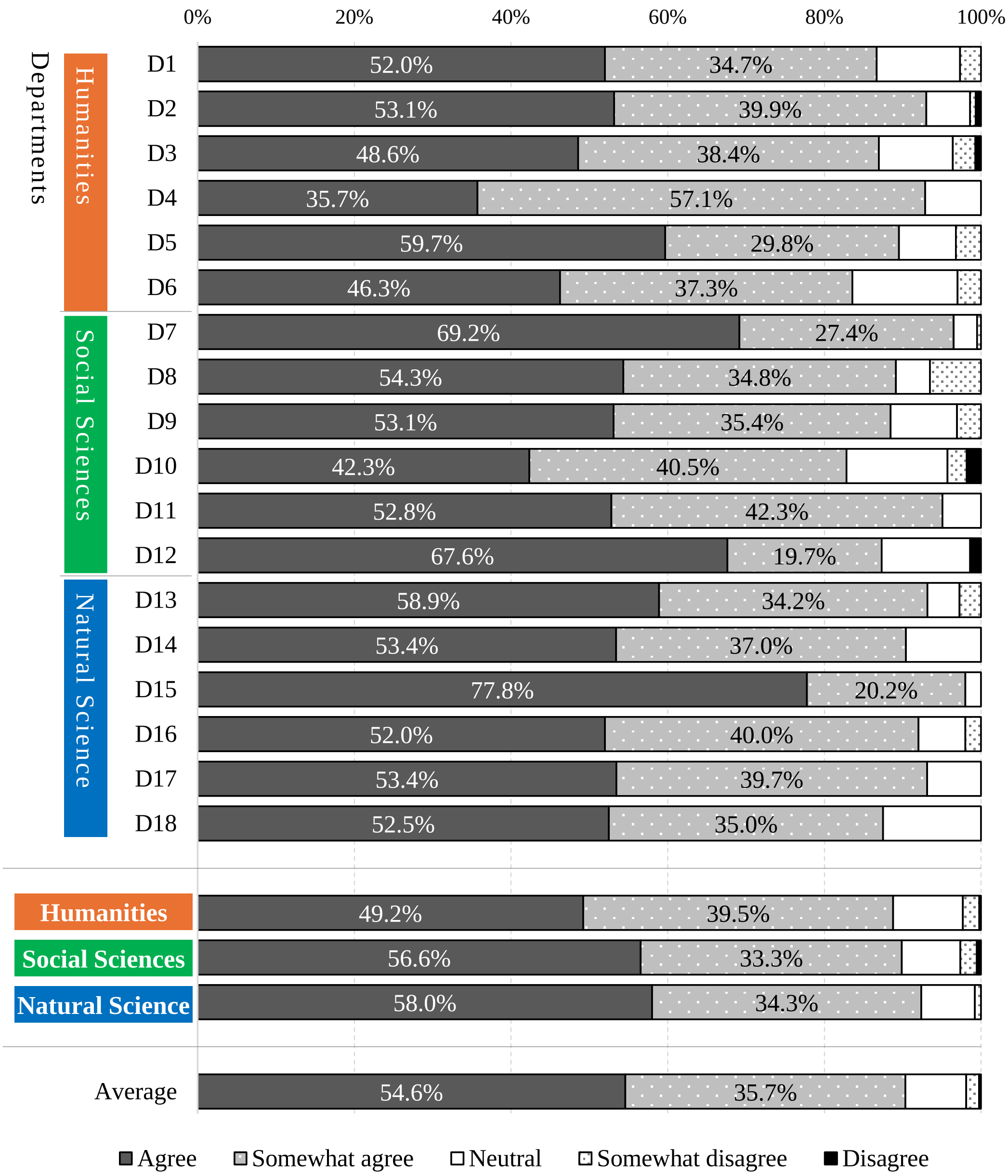

Figure 3: Learning Needs for Data Science in University Courses

Next, we asked whether the students intended to enroll in data science-related courses at present (Table 3).

Table 3: Percentage of Students Considering Enrollment in Each Data Science Subject.

		Basic Data Processing	Big Data Science
Departments	D1	10.7%	4.0%
	D2	20.1%	4.2%
	D3	19.6%	1.4%
	D4	14.0%	1.8%
	D5	13.6%	0.8%
	D6	13.2%	1.5%
	D7	9.4%	1.0%
	D8	8.7%	0.0%
	D9	14.6%	1.5%
	D10	14.6%	0.0%
	D11	23.8%	0.7%
	D12	26.4%	2.8%
	D13	28.4%	1.4%
	D14	39.7%	0.0%
	D15	38.4%	5.1%
	D16	35.3%	2.0%
	D17	43.1%	6.9%
	D18	33.8%	1.3%
Humanities		**17.6%**	**2.5%**
Social Sciences		**16.9%**	**1.0%**
Natural Science		**38.4%**	**2.9%**
Average		**24.3%**	**2.1%**

For Big Data Science, the values are consistently low across all disciplines. Although students express interest in and perceive the usefulness of studying data science and AI, many hesitate to enroll in courses when specific, less familiar keywords such as machine learning or modeling are listed.

On the other hand, while there are slightly more students considering enrollment in Basic Data Processing, which comprises classes structured around somewhat more familiar terms such as statistics and Excel, it falls far from half. Although many students felt that skills related to statistics and data handling were necessary in the future, as indicated in Table 2, few students contemplated enrolling. Furthermore, students in scientific disciplines considered enrolling at more than twice the rate of other students. This suggests that students who perceived themselves to have a higher aptitude for the subject were more inclined to consider enrollment.

In essence, there is a significant disparity between the perceived utility of data science and willingness to enroll in related courses. Students' motivations for enrollment likely involve various factors such as ease of credit acquisition, enhancement of their expertise, and academic performance for scholarships, among other considerations. Considering the nation's goal of fostering AI-ready citizens, the disadvantages arising from this situation are considerable. Despite recognizing the necessity of data science, most students graduate without learning it.

Based on the above, we argue that making the remaining two subjects of the data science curriculum mandatory along with Information Literacy is crucial within the CHS. In addition, there are advantages in terms of educational methodologies and content. At present, both subjects can be taken in any year and order, resulting in a diverse mix of students from different years and disciplines in the classroom with substantial disparities in knowledge and skill levels. Making both subjects mandatory will ensure that all students take them in the same year, leading to an expected homogenization of student levels compared to the current situation and making it easier to design education considering the sequence of subjects (Figure 4).

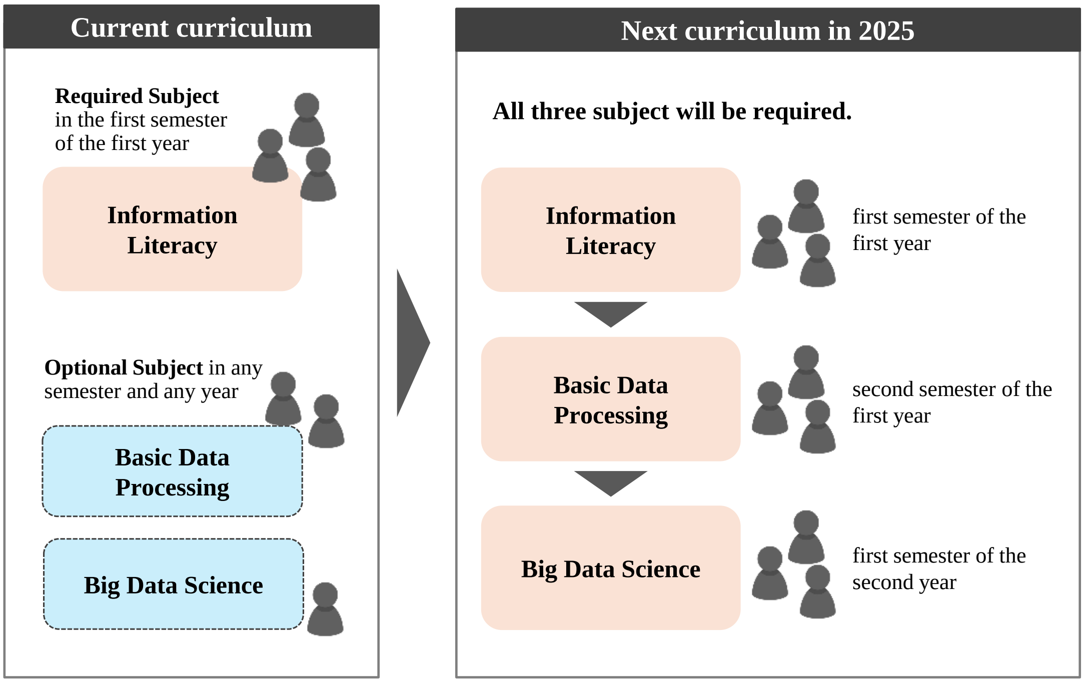

Figure 4: Comparison between the Current Curriculum and the Planned Curriculum.

4 Conclusion

In this study, we investigated students' perspectives on the increasing prominence of data science education each year, examining the extent to which they perceived its usefulness and their willingness to enroll in related courses. The results revealed that approximately 90% of the students, regardless of their field of study, recognized the importance of data science. However, while there is a general sense of the need for learning, there were slight variations across disciplines regarding specific skill sets such as ethics and data handling.

Furthermore, despite the high demand for data science among students, a considerable lack of awareness regarding enrollment in related courses was revealed. Students do not make enrollment decisions based solely on the perceived future utility of courses; various factors such as the importance of other subjects and the ease of obtaining good grades are also considered. Relying solely on student discretion would not suffice to achieve the goal of nurturing professionals capable of AI readiness. From this perspective, we believe that the significance of CHS's plan to make the three courses related to data science literacy mandatory in the future is crucial.

An important future task is to assess the consequences of mandating these courses. This is especially crucial in institutions such as ours, where STEM and non-STEM students mix and students from various fields study together. It is vital to determine whether it is suitable to have students from diverse academic backgrounds in the same classroom, being taught the same content and level using the same approach. We also need to evaluate the potential disparities in student performance and whether there is a risk of higher dropout or retention rates, particularly among those finding data science challenging. Continuous investigations and enhancements are essential to effectively address these concerns.

References

[1] Japanese Cabinet Office, "Society 5.0 - Science and Technology Policy - (in Japanese)," https://www8.cao.go.jp/cstp/society5_0/ (accessed 15-Apr-2024).

[2] Japanese Cabinet Office, "The Vision of a Hyper-Smart Society and Initiatives Toward a Hyper-Smart Society (in Japanese)," https://www8.cao.go.jp/cstp/tyousakai/kiban/3kai/siryo1.pdf (accessed 15-Apr-2024).

[3] Japanese Cabinet Office, "Human-centered AI Social Principles (in Japanese)," Integrated Innovation Strategy Promotion Council, https://www8.cao.go.jp/cstp/aigensoku.pdf (accessed 15-Apr-2024).

[4] Japanese Cabinet Office, "Overview of the AI strategy 2019 and Status of Initiatives (in Japanese)," Director-General for Policy Coordination, Cabinet Office, https://www5.cao.go.jp/keizai-shimon/kaigi/special/reform/wg7/20191101/shiryou1.pdf (accessed 15-Apr-2024).

[5] Ministry of Education, Culture, Sports, Science and Technology, "Approved Program for Mathematics, Data science and AI Smart Higher Education, designated (in Japanese)," https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00001.htm (accessed 15-Apr-2024).

[6] National Institute of Science and Technology, "Status of Students in Higher Education Institutions (in Japanese)," https://www.nistep.go.jp/sti_indicator/2022/RM318_32.html (accessed 15-Apr-2024).

[7] Ministry of Education, Culture, Sports, Science and Technology, "Publication of the 2021 Basic School Survey (in Japanese)," https://www.mext.go.jp/content/20211222-mxt_chousa01-000019664-1.pdf (accessed 15-Apr-2024).

[8] Google, "Google Trends," https://trends.google.co.jp/trends (accessed 15-Apr-2024).

[9] Davenport, T. H., Patil, D. J., "Data scientist". Harvard business review, Vol.90(5), pp.70-76, 2012.

[10] Peter Naur, "Concise Survey of Computer Methods," Petrocelli Books, 1974.

[11] Wu, C. F. J., "Statistics = Data Science?," in 7th series of P. C. Mahalanobis Memorial Lectures, 1997. http://www2.isye.gatech.edu/~jeffwpresentations/datascience.pdf (accessed 15-

Apr-2024).

[12] Wing, J. M., "The Data Life Cycle, " Harvard Data Science Review, Vol.1(1), 2019.

[13] American Statistical Association, "ASA Newsroom," https://www.amstat.org/ASA/Newsroom.aspx (accessed 15-Apr-2024).

[14] Pedersen, A.Y., Caviglia, F., "Data literacy as a compound competence". In T. Antipova & A. Rocha (Eds.), Digital science, Vol.850, pp.166–173, 2019.

[15] Cassel, L., & Topi, H. "Strengthening Data Science Education Through Collaboration. Alexandria", Virginia: National Science Foundation, 2016.

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

[17] Clancy, T. R., Bowles, K. H., Gelinas, L., Androwich, I., Delaney, C., Matney, S., Sensmeier, J., Warren, J., Welton, J., & Westra, B., "A call to action: Engage in big data science," Nursing Outlook, Vol.62(1), pp.64-65, 2014.

[18] Gold, M., McClarren, R., & Gaughan, C., "The lessons Oscar taught us: Data science and media & entertainment," Big Data, Vol.1(2), pp.105-109, 2013.

[19] J. Engel, "Statistical literacy for active citizenship: A call for data science education," Statistics Education Research Journal, Vol.16(1), pp.44-49, 2017.

[20] National Academies of Sciences, Engineering, and Medicine, "Data Science for Undergraduates: Opportunities and Options," Washington, DC: The National Academies Press, 2018.