Data-Driven Analysis of Skills Gaps in Emerging IT Fields: Aligning Pakistan's Education with Industry Demands for 2025 and Beyond

Mubashar Hussain MSDS24031
Rana Muhammad Sohaib Aasim MSDS22055
Umar Rafaqat MSDS24068
Muhammad Huzaifa Bhatti MSDS24038

***Abstract*—This study highlights the skill gaps in emerging IT fields by exposing the gaps between IT higher education and the IT industry of Pakistan. A diverse dataset was collected using survey methodology among the IT students and IT professionals across all over the academia and industry of Pakistan. The datasets from job posting websites and academic curriculum of Pakistan's higher education institutions were also collected to get comprehensive and in-depth insights to the matter, which would be ultimately actionable for bridging these gaps. Moreover, the paper provided a view on regional disparities by using the comprehensive data collection approach. Finally, both the qualitative and quantitative data provided the in-depth insights to deeply understand the reasons behind the skills gaps in emerging IT fields.**

I. Introduction

A. Overview

The IT sector is the spine of the modern age. It's been almost 100 years, and the field is evolving exponentially. Ultra-dynamic nature of the IT industry makes it hard for the education sector to stay parallel and it's even more difficult for under developing countries to match the growing demands of the IT industry. Moreover, the AI revolution is on the way to challenge the education sectors across the globe. Modern trends in computer science and emerging IT technologies imposed serious challenges to both IT academia and corresponding industry. Due to lack of resources and insufficient talent, it is hard for under developing countries to continuously update the academic curricula and learning methodologies with the evolving pace of industry. In short, the emerging IT skills and education discrepancies in Pakistan need innovative and context-specific training solutions to stabilize the economy of the country as this sector is the most crucial factor in economic growth of the country nowadays. The exact reasons behind the flaws of the education system of Pakistan are numerous and need a comprehensive analytical study, here we will just provide the data driven insights on these gaps between IT academia and industry of Pakistan, which would be helpful in taking decisions to bridge the gaps between IT academia and the corresponding Industry of the Pakistan. Moreover, this study will highlight the regional disparities in emerging IT skills across all over the country. This would also be helpful for policymakers to take corrective measures in the education sector of Pakistan to stabilize and boost-up the economy throughout the country.

B. Problem Statement

A significant gap in emerging IT fields between IT academia and corresponding industry of Pakistan has been developed in previous decade due the rapid advancement in the field of artificial intelligence and other modern trends, causing the mis-match in skills being taught and industry demands.

C. Research Objective

This study aims to identify the emerging IT skills gaps using a data-driven statistical approach and find the in-depth actionable insights to propose precise corrective measures for bridging the gaps between IT academia and industry that will help fulfilling the future demands of industry mainly attributed to artificial intelligence.

D. Challenges

The main challenge in this research project is the data collection, which includes:

1) Primary data i.e. survey data from IT graduates, professionals and educators across all over Pakistan.
2) Secondary data collection from the job posting platforms using data scrapping technique.
3) Curricula of various universities and institutes across all regions of Pakistan.

II. Literature Review

Abelha et al.'s 2020 [1] done a systematic review on graduate employability that reveals a global skills mismatch between higher education and employer's needs and demands, particularly relevant to IT. Despite efforts, universities often fall short in preparing graduates for handling the challenges of the job market.

Li et al.'s 2021 [2] study on data science skills in manufacturing offers insights applicable to Pakistan's IT sector. It highlights crucial technical skills like data management, machine learning, and automation, relevant across industries. The research exposes significant mis-match between employer's demands and job seeker's skills, mirroring potential issues in Pakistan's IT education.

Zaman and Nadeem's 2024 [3] comparative study of skill-oriented education in Pakistan and India reveals contrasting approaches: Pakistan emphasizes life skills, while India focuses on career development. Both countries face implementation issues due to infrastructure problems and outdated curricula, with Pakistan experiencing more significant difficulties.

A paper on bridging the skills gap in robotics [4] explored the rapidly increasing demand for robotics professionals but there is shortage of the professionals with required skill sets. It highlighted that skills in programming and software-specific knowledge are most important but it must be integrated with practical skills such as assembly and welding. The diversity in application areas of robotics created demand for multi-disciplinary skills such as artificial intelligence, new materials.

The research paper [5] conducted a study to examine the disparity in digital and AI skills across different regions of Europe. It highlights a notable difference in levels of skills between the residents of rural or small community areas and large urban areas. This discrepancy is attributed to un-even opportunities to access the digital-infrastructure and learning resources.

The research paper [6] presented a novel approach and procedural methodology for analyzing and comparing the academic curricula and their corresponding skill sets demanded in industry. As there's trend of rapid advancement in digital technologies such as automation, this caused many institutions lagging in providing the up-to-date skill sets to meet the dynamic demands of the growing industry.

This Sweden based research [7] uses a text-analysis tool called JMAR to compare the important keywords in syllabi of higher education (HEI) and job-postings. It focuses on three main research questions. Technologies taught in HEI, demands for technologies in job-postings and limitations of the tool JMAR used in this research. They concluded that where HEI teaches most of the technologies required in job postings it only does so at a broader conceptual level.

The research paper [8] conducted a systematic search using PRISMA guidelines to study 40 different papers for understanding and measuring skill gaps in the industry 4.0. Most of the papers used the quantitative approach i.e. surveys. The paper concluded that the gap between competency of employees and demands of the employer is the most common skill gap mentioned in almost all the papers.

The research paper [9] presents a bibliometric analysis of 10,214 scholarly articles from 2010-2021 for the purpose of aligning university education with job market demands. To reveal growing areas of technical skills and soft skills for the industry 4.0 and 5.0. The study used Scopus data and employed co-authorship, citation, and keyword co-occurrence networks.

This research paper [10] conducted a systematic literature review (SLR) and meta-analysis of 35 research papers which gathered over 4,000 data points from various sources like surveys, job advertisements and interviews. The findings of this study imply revising the SE course curriculum to focus on SE skills such as testing, project management, design and some soft skills like leadership, teamwork and critical thinking.

This research paper [11] uses qualitative methods such as interviews with HR's, focus groups with students and analysis of curriculum for highlighting the fact that formal education institutes mostly focus on technical skills and ignore emotional and social skills. Due to this IT companies often fail to find well-rounded professionals which leads them to do extra training and its cost.

Paper [12] used data collected from 659 software practitioners across 14 different countries is focused on bridging the gap between industry and academics and calls for more Industry-Academic Collaboration (IAC). Many practitioners had not taken courses on essential topics like Software Configuration Management or Software/Systems Analysis, despite their importances and frequent use in the industry.

Paper [13] which was conducted in Kosovo used two methods of data collection, one is interviews and the other is surveys. According to the employees' main criteria for their selection were certifications from online platforms and certifications from training centers.

Paper [14] provided a study based on a survey-specific quantitative approach that explored the impact of AI on industries in emerging economies. It highlighted significant skill gaps in AI integration, machine learning and ethical considerations. Nowadays AI is rapidly integrating with the industries, especially IT industry but many underdeveloped economies still lag in providing up-to-date knowledge and adequate context-oriented training.

Paper [15] uses different sources such as government reports, industry studies and surveys to collect data to understand the demand for digital and STEM skills. Due to rapid economic growth and COVID-19 pandemic, there is a shortage of skilled workers. Reforms in educational systems are needed to increase STEM enrolment and increase skills on an individual level.

Papers [16] draws from case studies, practical experiences and literature to find better ways which educational institutes can adopt for student's growth during a pandemic. Real-time response systems, learning management systems and analytics-enabled platforms help personalize learning and provide better learnings in remote environments.

And finally, the paper [17] conducts interviews with 13 IT professionals to find the skill gaps in the post-pandemic market. The result of the interview concludes that soft skills such as communication, teamwork are lacking in the post-pandemic market. To remedy this, educational institutes need to develop better soft skills along with technical skills in their students.

III. Data Collection

We collected mainly three datasets to use for the statistical analysis.

A. Survey

Survey data is the main dataset we used to conduct this research. The questionnaire for this survey is designed to be

bias-free and diverse to make this data generalizable for getting deeper insights. The audience of this survey were IT students, fresh graduates, employees, employers and educators from all over the Pakistan.

TABLE I
RESPONDENTS' CURRENT ROLES SUMMARY

Respondents' Current Role	Count	Percentage (%)
Employed in Emerging IT Field	69	56.56
Final-Year Undergrad Student of IT Field	20	16.39
Graduated or Final-Year Grad Student of IT Field	24	19.67
Faculty Member in IT Field	8	6.56
Employer/Hiring Manager in Emerging IT Field	1	0.82

Survey Insights

We surveyed a sample of size 122, the count in the Figure 1 below gives the insights that a great number of students are not able to learn the core emerging skills from their respective universities.

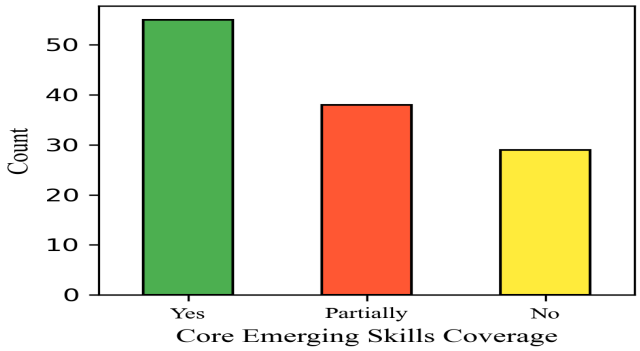

Fig. 1. Coverage of Core Emerging Skills

Moreover, many universities of Pakistan are not providing the required additional certifications to their enrolled students to cover the missing skills as shown in the Figure 2 below.

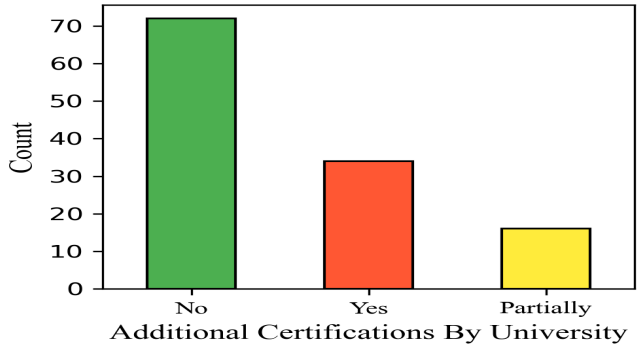

Fig. 2. Additional Certifications By Universities – Count

The question "What core emerging IT skills are missing in your degree?" provided us insights regarding the emerging IT skills required in industry which were not taught to the IT students in their respective degrees. Below is the exact state of the missing skills.

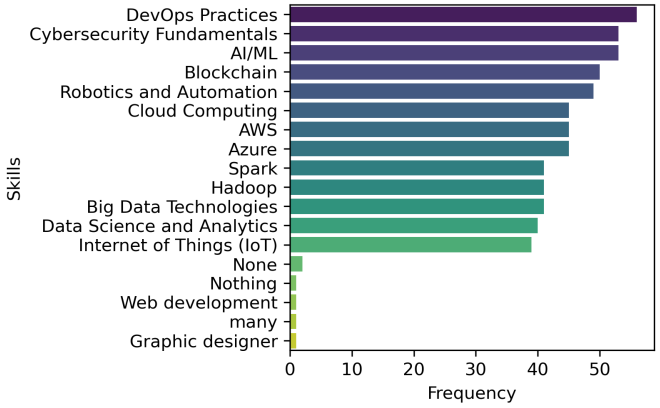

Fig. 3. Missing Emerging IT Skills in Degrees

The high frequency of missing core emerging IT skills as shown in Figure 3 suggests that higher educational institutions may be lagging in regularly updating their curriculum and teaching standards according to industry's technological and workforce needs.

One of the interesting insights in the survey dataset tells us about regional discrepancies in terms of adequate opportunities and availability of resources as shown in the following figure.

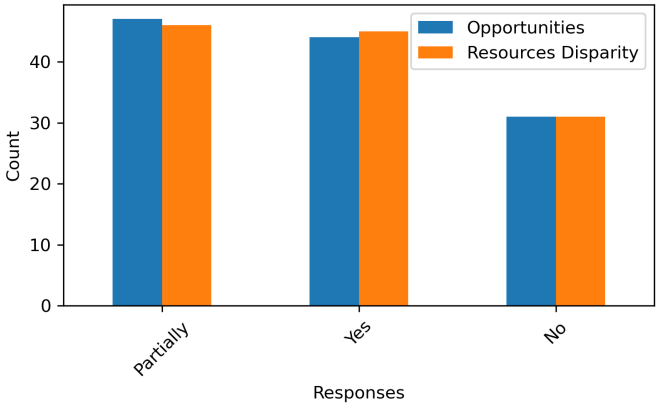

Fig. 4. Adequate Opportunities and Resources Disparity

As shown in the Figure 4, responses on regional opportunities and disparities were almost equal for each option. Those who said that they were given adequate opportunities partially were also almost partial on the disparity of resources for making full of those opportunities. This trend repeats itself for Yes and No responses as well.

B. Curricula

For our curriculum dataset, we focused only on IT programs that universities in Pakistan are offering. We selected 18 uni-

versities across all over the Pakistan with varying excellence and gathered their curriculum through their official websites. Below is the summary of this dataset.

TABLE II
SUMMARY OF CURRICULUM DATASET

Mark	Count
Universities	18
Programs/Degrees	9
Skills/Fields	18

Curriculum Insights

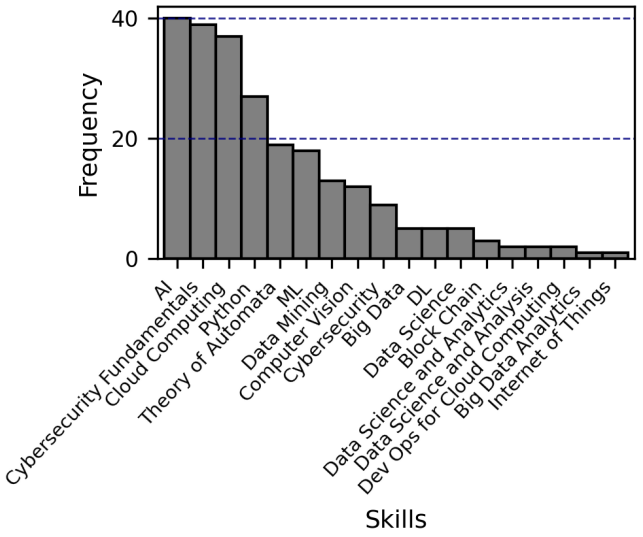

Fig. 5. Frequency of Courses in IT Degrees

In the curriculum dataset, we got the fact that almost all the Pakistani universities offering the courses on core emerging skills such as AI/ML, cybersecurity, cloud computing etc., on the other hand, it is known that certain potion of emerging IT domain is still missing such as blockchain, big data etc.

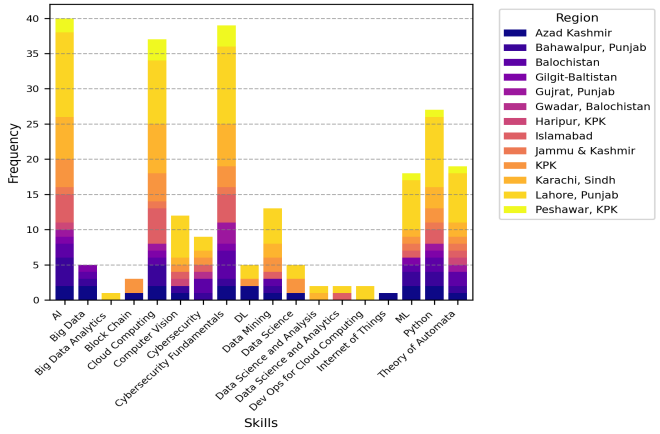

Fig. 6. Emerging IT Courses Exposure in Different Regions

There are certain emerging IT skills i.e., AI, cloud computing and cybersecurity that are being covered by almost all the regions of Pakistan, but on the other hand, the region Gwadar and Haripur have limited exposure to the emerging IT skills.

C. Job Postings

To collect jobs data from the IT sector in Pakistan, we utilized a browser extension called "Indeed Scraper". While gathering data, we removed any irrelevant job postings, such as those outside the IT sector, and removed duplicates or incomplete entries. This step helped us to ensure that the dataset accurately reflected the range of emerging IT job opportunities available on Indeed.

TABLE III
SUMMARY OF JOB POSTING DATASET

Mark	Insight
No. Of Job Postings	33
Most Frequent Location	Lahore
Most Frequent Job-Type	Full Time
Most Demanding Skills	Python
Popular Job Headings	Data Science, DevOps Engineer

The dataset was providing scattered information. Particular job skills were mentioned in it, as we needed to analyze it alongside the curriculum and missing skills mentioned in the survey, so we decided to group sub-skills under the main skills at broader level, such as python libraries into a larger section of python. Table IV is showing the cleaned data that is used in statistical testing.

TABLE IV
SUMMARY OF CLEANED DATA FROM JOB POSTINGS

Skill	Count
Python	29
Cloud Computing	19
Artificial Intelligence	26
Big Data	9
Data Analytics	40
Cyber Security	4
DevOps	3

Job Posting Insights

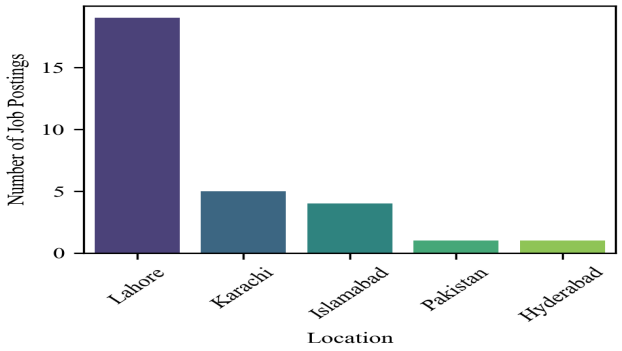

Fig. 7. Job Distribution by Location

One the interesting facts of IT jobs is its concentration in the Lahore city. In our dataset of job postings, most of the jobs were posted in this city of Punjab, Pakistan.

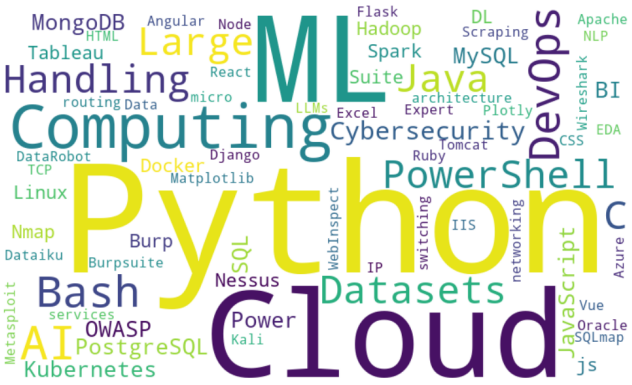

Fig. 8. Frequently used Words in Job Postings

The role of python programming in emerging IT fields such as AI/ML and data science and analytics is very prominent. It is most demanded core skill which is required in the area of emerging IT fields. Other most frequently asked core skills in emerging IT roles include cloud computing, data science, machine learning and devOps.

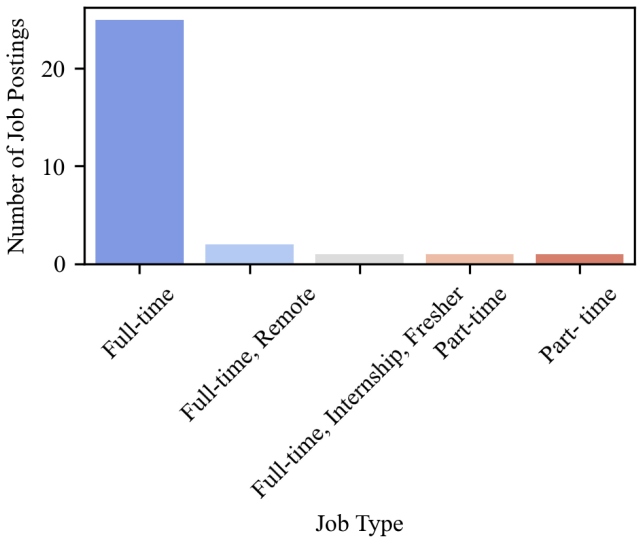

Fig. 9. Job Distribution by Job-Type

It is seen that remote and part time jobs are very less common as compared to the frequently occurring on-site full-time jobs in Pakistan.

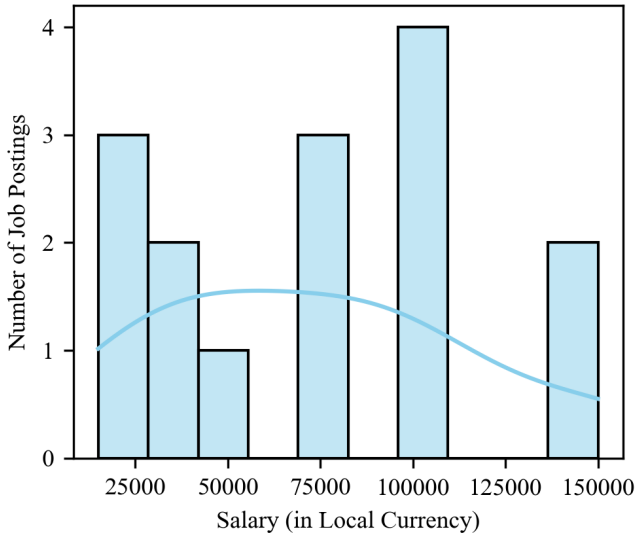

Fig. 10. Job Distribution by Salary

Most of the emerging IT roles were offering almost a hundred thousand PKR salary to the employees as shown in the Figure 10.

IV. Statistical Analysis

A. Hypothesis Testing

For our hypothesis testing, we created mainly three hypothesis on the basis of the multiple kinds of datasets we collected.

1) Hypothesis 1: Firstly, we did analysis on the skills in curriculum and skills in job postings. As our goal is to find dependencies between curriculum skills and job posting skills, that's why we used Chi-squared test.

The Chi-squared statistic is given by the formula:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Where:

- O_{ij} = The observed frequencies.
- E_{ij} = The expected frequencies.

We used python programming as a tool for computations. Following are the null and alternative hypothesis:

Ho : There is no association between skills that are mentioned in curriculum and job postings.

H1 : There is an association between skills that are motioned in curriculum and job postings.

This analysis tells us whether the courses, universities offer in their IT programs can give students the required skills demanded by the IT industry.

2) Hypothesis 2: Secondly, we did analysis on the missing skills mentioned in surveys and skills required for job postings. The particulars of this analysis were on the basis that graduates told us about what skills their degree did not teach them which they should have and whether they are required for job postings or not. As we needed to find dependencies between missing skills and job posting skills, we used the Chi-squared test with the following null and alternative hypothesis,

Ho: There is no association between missing skills mentioned in the survey and job postings.

H1: There is an association between missing skills mentioned in the survey and job postings.

This analysis will tell us whether graduates were able to learn the skills from their respective university's IT programs, demanded by the IT industry.

3) Hypothesis 3: Finally, we applied statistical testing on highest IT degree and rate of job preparation at the end of degree. For this we divided highest IT degrees into BS and MS, and separated each one's ratings. Then to find if there is an association between them, we used chi-squared test with the following null and alternative hypothesis,

Ho: There is no association between Highest IT Degree (BS and MS) mentioned in the survey and job preparation ratings.

H1: There is an association between Highest IT Degree (BS and MS) mentioned in the survey and job preparation ratings.

This analysis will tell us whether job preparation ratings depend upon the type of highest degree.

B. Hypothesis Testing Results

1) Hypothesis 1 - Result: The contingency table for the testing is given below:

TABLE V
CONTINGENCY TABLE FOR HYPOTHESIS-1

	AI	**Data Analytics**	**Cloud Computing**	**Cyber Security**	**Python**	**Dev Ops**	**Big Data**
Curriculum	58	22	38	49	27	2	6
Job Posting	26	40	19	4	29	3	9

After using python programming to apply chi-squared on this table, the following values are computed.

Chi-squared statistic: 49.544
Degree of freedom: 6
P-value: 5.8012e-09

As the p-value is less than significance level which is 0.05. We can reject our null hypothesis. This means our alternative hypothesis which suggests "There is an association between curriculum skills and job posting skills." is proven correct. This suggests that the curriculum of most of the universities can give students the required skills to fulfill the industry needs.

2) Hypothesis 2 - Result: The contingency table of our testing is given below:

TABLE VI
CONTINGENCY TABLE FOR HYPOTHESIS-2

	Dev Ops	**Cyber Security**	**AI**	**Block chain**	**Robotics and Automation**	**Cloud Computing**	**Big Data**	**Data Science**	**IoT**	**Python**
Missing Skills	56	53	53	50	49	45	41	40	39	0
Job Postings	3	4	26	0	0	19	9	40	0	29

This statistical testing provided the following results:

Chi-squared statistic: 194.5644
Degree of freedom: 9
P-value: 4.5611e-37

As the p-value is less than the significance level which is 0.05. We reject our null hypothesis. This suggests that a lot of graduates feel that universities are failed to deliver these core emerging IT skills for some reasons.

3) Hypothesis 3 - Result: The contingency table of our testing is as follows:

TABLE VII
CONTINGENCY TABLE FOR HYPOTHESIS-3

	1	**2**	**3**	**4**	**5**
B.S	13	25	29	8	8
M.S	5	2	9	6	0

Following values are computed,

Chi-squared statistic: 9.84473
Degrees of freedom: 4
P-value: 0.043

As our p-value is less than significance level which is 0.05 we reject our null hypothesis. This means that the respondent's highest IT degree affected how they rated job preparation at end of the degree.

V. CONCLUSION

There seems a significant skills gap in emerging IT fields in Pakistan provided in the survey dataset, which is assumed as the direct statement of the relevant people. Moreover, a statistical test proved that the core emerging skills missing in the degrees that were mentioned by the IT people in the survey dataset, are similar to those of demanded in the job postings by the corresponding IT industry. This implied that many universities of Pakistan are failed to deliver the industry-oriented emerging IT skills to the students. However,

a statistical test also inferred that the curriculum of most of the higher education institutions of Pakistan includes the core emerging IT skills that are demanded in the job postings in emerging IT fields. This concluded that the relevant courses are offered by most of the universities to the students, but on the basis of survey data insights, we reached to the point that there's gap in teaching standards, course contents and availability of quality resources for robust learning of students.

VI. Future Work Direction

This analysis is clearly mirroring the direction for future research work. There is need to do a robust analysis of the course contents being taught by the universities of Pakistan, especially in the domain of IT programs. Moreover, the teaching standards and provision of quality resources to the students is also needed to be analyzed for better understanding the emerging IT skill gaps mentioned by a great number of IT people.

References

[1] D. M. F. S. Marta Abelha, Sandra Fernandes and A. T. Ferreira-Oliveira, "Graduate employability and competence development in higher education — a systematic literature review using prisma," *Published in Sustainability*, vol. 12, no. 4, pp. 1–23, 2020. [Online]. Available: http://dx.doi.org/10.3390/su12155900

[2] S. K. M. M. X. J. Guoyan Li, Chenxi Yuan, "Data science skills and domain knowledge requirements in the manufacturing industry: A gap analysis," *Published in Journal of Manufacturing Systems*, vol. 60, pp. 692–706, 2021. [Online]. Available: https://doi.org/10.1016/j.jmsy.2021.07.007

[3] D. M. A. N. Muhammad Rafiq-Uz-Zaman, "Comparative analysis of skill-based education curriculum in pakistan and india: A contemporary review," *In Journal of Academic Research for Humanities*, vol. 04, no. 3, pp. 187–197, 2024. [Online]. Available: https://zenodo.org/records/13646448

[4] N. Shmatko1 and G. Volkova, "Bridging the skill gap in robotics: Global and national environment," *SAGE Journal*, pp. 1–13, 2020.

[5] T. S. T. M. Long Pham, Barry O'Sullivan, "Addressing digital and ai skills gaps in european living areas: A comparative analysis of small and large communities," *Published in Proceedings of the AAAI Conference on Artificial Intelligence The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)*, 2024. [Online]. Available: https://doi.org/10.1609/aaai.v38i21.30357

[6] R. M. T. F. Föll, Patrick, "Bridging the gap between industry skill demand and university skill provision," *Twenty-Ninth European Conference on Information Systems (ECIS 2021), Marrakesh, Morocco.*, 2021.

[7] L.-M. A. A. Felix Dobslaw, Kristian Angelin, "The gap between higher education and the software industry — a case study on technology differences," *European Conference on Software Engineering Education (ECSEE 2023)*, p. 11, 2023. [Online]. Available: https://doi.org/10.1145/3593663.3593690

[8] M. J. J. S. R. H. Pauliina Rikala, Greta Braun, "Understanding and measuring skill gaps in industry 4.0 — a review," *Published in Technological Forecasting Social Change*, 2024. [Online]. Available: https://doi.org/10.1016/j.techfore.2024.123206

[9] A. O. K. S.-U. H. Naif Radi Aljohani, Ahtisham Aslam, "Bridging the skill gap between the acquired university curriculum and the requirements of the job market: A data-driven analysis of scientific literature," *Published in Journal of Innovation Knowledge*, vol. 7, July–September 2022. [Online]. Available: https://doi.org/10.1016/j.jik.2022.100190

[10] E. T. C. C. M. F. Vahid Garousi, Görkem Giray, "Aligning software engineering education with industrial needs: A meta-analysis," *Published in Journal of Systems and Software*, 2019. [Online]. Available: https://doi.org/10.1016/j.jss.2019.06.044

[11] L. O. C. Vera G. Goulart, Lara Bartocci Liboni, "Balancing skills in the digital transformation era: The future of jobs and the role of higher education," *Industry and Higher Education*, pp. 1–10, 2021.

[12] D. AKDUR, "A survey on bridging the gap between software industry and academia: Preliminary results," 2019. [Online]. Available: https://www.researchgate.net/publication/334559697

[13] E. Ahmeti, "Skills gap in the information technology sector in kosovo," *Thesis.Rochester Institute of Technology.Accessed*, 2022. [Online]. Available: https://repository.rit.edu/theses

[14] N. T. A. S. A. F. C. M. R. Gursahildeep Singh Sidhu, Md Abu Sayem, "Ai and workforce development: a comparative analysis of skill gaps and training needs in emerging economies," *International journal of business and management sciences*, vol. 04, pp. 12–28, 2024. [Online]. Available: https://doi.org/10.55640/ijbms-04-08-03

[15] P. Mahboubi, "The knowledge gap: Canada faces a shortage in digital and stem skills," *Commentary 626. Toronto: C.D. Howe Institute*, 2022.

[16] I. O. Pappas and M. N. Giannakos, "Rethinking learning design in it education during a pandemic," *Frontiers in Education*, vol. 6, 2021.

[17] N. R. N. K. Udanee Sachithra Samarasinghe, Chathurini Kumarapperuma, "It graduate employability skills: Before and after covid-19 from the perspective of it experts (through the eyes of it experts)," 2022. [Online]. Available: https://www.researchgate.net/publication/361902363