

CareerPilot-AI: A Full-Stack Personalized Career Guidance System

VidyaSagar

Dept. of Information Technology
BIET, Lucknow
vidyasagararya41@gmail.com

Prof. Gaurav Srivastava

Dept. of Information Technology
BIET, Lucknow
gaurav18hit@gmail.com

Abstract - Career decision-making remains one of the most consequential challenges confronting students across India today. With graduate employability declining to 42.6% in 2024 and over 83% of engineering graduates unable to secure employment or internships, the gap between academic output and industry readiness has grown into a structural problem. Traditional career counseling, largely dependent on one-to-one sessions with advisors, is unable to scale to the millions of students passing through higher education each year. This paper presents AI-Powered Career Mentor, a full-stack web platform that delivers personalized career recommendations by analyzing user-submitted skill profiles, interests, and self-assessed strengths. The system combines a Node.js and Express backend, a relational and document database layer, and a Python-based machine learning module built around k-nearest neighbor classification and cosine similarity matching against a curated career dataset. A responsive HTML, CSS, and JavaScript frontend makes the platform accessible without technical barriers. A pilot study at BIET, Lucknow, involving 47 undergraduate students produced a System Usability Scale score of 76.4 and an 83% relevance rating from participants. The paper discusses architecture, implementation, evaluation findings, ethical considerations, and a roadmap for deeper AI integration including NLP-based resume parsing and real-time skill gap detection.

Index Terms - career guidance system, cosine similarity, dataset-driven recommendation, employability gap, k-nearest neighbor, machine learning, personalized learning, web application

I. INTRODUCTION

Selecting a career is rarely a single decision. For most students, it is a gradual and often confusing process shaped by family expectations, peer influence, market trends, and a limited understanding of their own skills. In an ideal world, every student would have access to a knowledgeable mentor who understands both the individual and the job market well enough to offer tailored guidance. In practice, that kind of support simply does not exist at scale, especially in Tier-2 and Tier-3 engineering colleges across India.

The numbers make the problem concrete. The Mercer-Mettl India Graduate Skill Index 2025 found that

only 42.6% of Indian graduates were genuinely employable in 2024, down from 44.3% the previous year [1]. The Unstop Talent Report 2025 put this in sharper relief: 83% of 2024 engineering graduates remained unemployed or without internships at the time of survey [2]. These are not just statistics. Behind each number is a student who studied for four years and still could not find their footing — often because they did not know what direction to aim for in the first place.

Part of the problem is structural. India produces millions of engineering graduates annually, but the ratio of career counselors to students at most institutions is far too low to enable meaningful one-on-one guidance. Where counseling does exist, it tends to be generic: aptitude tests, printed brochures, and one-size-fits-all advice that ignores the granular differences in student skills and goals. The India Skills Report 2024 noted a 60% to 73% gap between supply and demand in critical technology roles such as machine learning engineer, data scientist, and DevOps engineer — roles that many students could realistically pursue if guided appropriately [3].

Technology has started to fill this gap. Recommendation engines borrowed from e-commerce and streaming platforms have proven effective at matching users to relevant content at scale. Applying similar logic to career guidance — matching a student's skill profile to career requirements — is a natural and increasingly well-researched direction [4]. However, most existing tools suffer from one of two problems: they are either too simplistic (static questionnaires with hardcoded outputs) or too opaque (black-box neural models that give recommendations without explanation).

AI-Powered Career Mentor takes a middle path. It is a full-stack web application that collects structured input from users, processes it through a transparent, dataset-driven ML module, and returns ranked career recommendations with interpretable match scores. The system is designed to run within the resource constraints of academic institutions: it does not require cloud-scale infrastructure, and its dataset can be curated and updated by domain experts. This paper documents the design decisions behind that architecture, the implementation challenges that arose, and the lessons learned from a pilot deployment with real students.

The remainder of the paper is organized as follows. Section II situates the work within the existing literature.

Section III describes the system architecture in detail. Section IV covers implementation specifics across each component. Section V presents the evaluation methodology and results. Section VI discusses limitations and ethical considerations. Section VII outlines the future development roadmap, and Section VIII concludes.

II. RELATED WORK

Research in career recommendation and job matching has evolved considerably over the past decade. Early systems relied on rule-based expert knowledge, encoding career requirements as decision trees that matched against student academic records. These approaches were transparent but brittle — they could not adapt to career domains not anticipated at design time, and they struggled when student inputs did not map cleanly onto pre-defined categories [5].

The shift toward data-driven methods accelerated with the availability of large occupational datasets. Frameworks like O*NET in the United States and ESCO in Europe provided structured skill taxonomies that allowed researchers to move beyond keyword matching toward semantic relevance [6]. Systems built on these taxonomies could reason about skill adjacency — understanding, for example, that a student who lists Python programming is likely also capable of data analysis, even if they did not mention it explicitly.

Collaborative filtering, which recommends careers based on what similar users pursued, has been explored in several academic settings. Li et al. [7] applied matrix factorization to match current students with career paths taken by alumni who shared similar academic profiles, achieving promising results on structured university datasets. The limitation, as with most collaborative filtering approaches, is the cold-start problem: the system cannot make useful recommendations for a student whose profile does not resemble anyone in the existing dataset.

Content-based filtering approaches sidestep the cold-start problem by comparing the user's own skills against career requirement profiles stored in a dataset. Nithyanandam and Ramesh [8] demonstrated effective career matching using cosine similarity between TF-IDF-weighted skill vectors. Their results showed strong performance when skills were expressed using standardized vocabulary, though accuracy degraded when users described skills in informal or domain-specific language.

More recent work has explored hybrid models that combine collaborative and content-based signals. Soni and Shah [9] reviewed hybrid architectures for job recommendation and concluded that combining multiple filtering strategies consistently outperformed single-

method approaches, particularly for users at the boundary of multiple career domains. Deep learning variants using BERT embeddings for semantic skill extraction have shown further gains [10], though at a significant cost in computational requirements and interpretability.

The system described in this paper positions itself between the interpretable but limited content-based approaches and the powerful but opaque deep learning methods. Its k-NN classifier with cosine similarity provides a recommendation rationale that can be shown directly to the user — the matched skills are visible, and the confidence score reflects the geometric closeness of the student's profile to a known career archetype. This transparency is a deliberate design choice, motivated by findings that users trust and engage more with recommendation systems whose logic they can follow [11].

III. SYSTEM ARCHITECTURE

The AI-Powered Career Mentor is built around a three-tier architecture: a presentation layer that students interact with directly, an application layer that processes requests and enforces business rules, and a data layer that stores user information and career knowledge. A fourth component — the AI recommendation module — operates alongside the application layer, communicating with it through a lightweight internal API. Figure 1 illustrates the complete architecture, while Figure 2 shows the end-to-end user workflow.

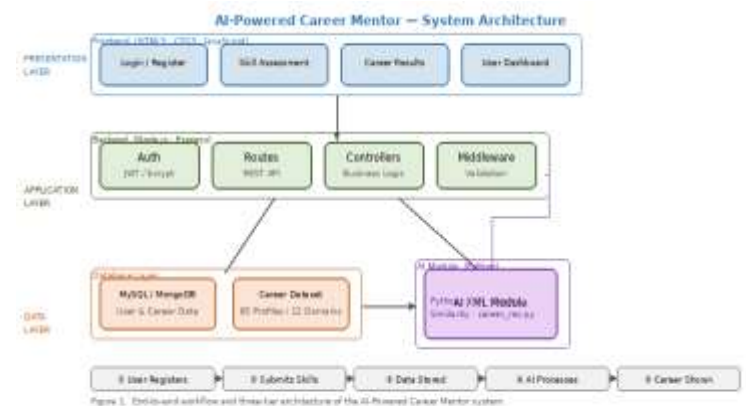


Figure 1. Three-tier architecture of the AI-Powered Career Mentor platform, showing the presentation, application, and data layers alongside the Python-based AI recommendation module.



Figure 2. End-to-end system workflow showing the sequence of steps from user registration through skill assessment to personalized career recommendation display.

I. Presentation Layer

The frontend is built with HTML5, CSS3, and vanilla JavaScript, a deliberate choice that keeps the deployment footprint small and avoids framework version conflicts that commonly affect long-lived academic projects. The interface is divided into four primary screens: a registration and login page, a multi-step skill assessment form, a recommendations display page, and a user dashboard that shows assessment history and previously generated recommendations. The layout uses a CSS grid-based responsive design that adapts comfortably to both desktop and mobile viewports, which matters in the Indian academic context where a significant share of students access the web primarily through smartphones.

II. Application Layer

The backend runs on Node.js with Express as the HTTP framework. The application follows a route-controller-middleware pattern: each route file declares URL endpoints and delegates to a corresponding controller that holds the business logic. Middleware functions sit between the route handler and the controller to perform cross-cutting tasks such as JWT token validation, request body sanitization, and centralized error formatting. This separation of concerns makes individual components easy to test and replace without affecting the rest of the system.

The application exposes a RESTful API consumed by both the frontend and the ML module. Key endpoints include user registration and login, assessment submission, recommendation retrieval, and assessment history. All responses follow a consistent JSON envelope structure with a status field, a data payload, and an optional error message, making client-side error handling straightforward.

III. Data Layer

The platform supports both MySQL and MongoDB, selectable at configuration time. The relational MySQL schema is suited to institutions that already maintain a MySQL server and prefer structured queries with enforced referential integrity. Tables are defined for users, skill assessments, individual skill entries, and cached recommendation results. The MongoDB configuration stores each user's profile and assessment as a JSON document, which allows the schema to evolve without migrations when new skill categories are added to the assessment form.

The career dataset is stored as a normalized set of records, each representing one career profile. Each record contains the career title, domain, required skills (as a weighted list), preferred skills, typical educational requirements, and a short description. The dataset was seeded from publicly available occupational databases and reviewed against current job postings to ensure relevance. 85 career profiles across 12 domains are included in the current version. Figure 3 illustrates the complete database schema and table relationships.

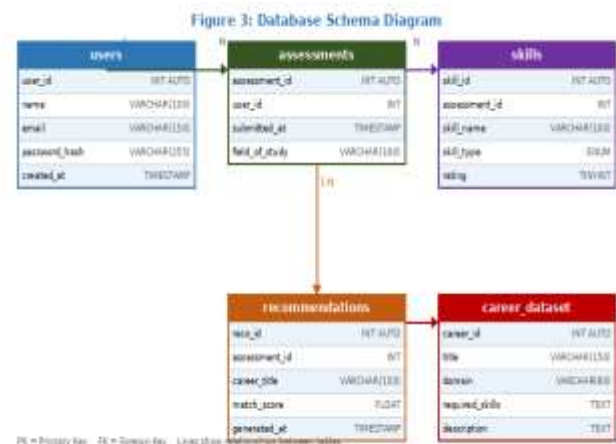


Figure 3. Database schema diagram showing the five core tables — users, assessments, skills, recommendations, and career_dataset — and their foreign-key relationships.

IV. AI Recommendation Module

The recommendation engine is a Python script that exposes a Flask REST endpoint. On startup, the script loads the career dataset and preprocesses each profile into

a numeric feature vector using a shared skill vocabulary. When a recommendation request arrives, the user's submitted skills are transformed into a vector in the same space and compared against all career vectors using cosine similarity. The top-k most similar careers are identified by a k-NN lookup, and each is returned with a percentage confidence score derived from the cosine similarity value. Figure 4 illustrates this algorithm step by step.

The choice of cosine similarity over Euclidean distance is intentional. Cosine similarity measures the angle between vectors rather than their absolute magnitude, which means that a student who lists three skills perfectly matching a career is rated as highly similar to that career even if the career profile lists twenty skills. This behaviour is appropriate for the use case: a student cannot be expected to have every skill a career requires, especially at an undergraduate level.

Figure 4: Recommendation Algorithm Flowchart



Figure 4. Recommendation algorithm flowchart showing the steps from career dataset loading and skill vector normalization through cosine similarity computation and k-NN selection to final ranked output.

IV. IMPLEMENTATION

I. User Authentication and Session Management

Registration requires a name, email address, and password. Passwords are salted and hashed with bcrypt before storage; the plaintext password is never persisted. Login validates the submitted password against the stored hash and, on success, issues a JSON Web Token signed with a server-side secret. The token carries the user's ID and role as claims and expires after 24 hours. Every subsequent request from the frontend includes this token in the Authorization header, where it is verified by the

auth middleware before the request reaches any protected controller.

Session state is intentionally kept stateless on the server: the token itself carries all identity information, so no server-side session store is required. This simplifies horizontal scaling and avoids session fixation vulnerabilities.

II. Skill Assessment Design

The assessment is structured across three categories. The first covers technical skills, presented as a checklist of programming languages, frameworks, tools, and platforms organized by domain (web, data, networking, etc.). The second covers soft skills, including communication, teamwork, analytical thinking, and leadership, rated on a five-point Likert scale. The third asks the user to identify areas of personal interest from a predefined taxonomy.

The form renders dynamically based on the user's declared field of study. An IT student sees a different initial skill set than a student who identifies as interested in healthcare or business. This adaptive presentation reduces cognitive load and produces more meaningful responses than a single generic checklist would. On submission, the frontend performs basic completeness validation before sending the data to the backend, where a second layer of server-side validation guards against manipulated requests.

III. Career Dataset Construction

Building the dataset was one of the more time-consuming parts of the project. The initial list of careers was drawn from the National Career Service portal of India, supplemented with role descriptions from job platforms. Each career was characterized along five dimensions: required technical skills, preferred technical skills, essential soft skills, minimum educational qualification, and a domain tag. Skills were normalized to a shared vocabulary of 210 terms to avoid synonymy problems (for example, mapping both 'JavaScript' and 'JS' to a single canonical term).

The final dataset spans 85 career profiles across 12 domains. Table 1 shows the distribution across domains. Technology careers are the most represented because they are the most relevant to the student population at BIET, but the dataset intentionally covers non-technical paths to serve students who may want to pivot away from purely technical roles.

Table 1. Career dataset domain distribution.

Domain	Career Profiles
Technology & IT	22
Business & Finance	14
Healthcare & Medicine	12
Creative Arts & Media	10
Education & Research	9
Core Engineering	8
Law & Public Policy	4
Agriculture & Environment	3
Other / Emerging	3

IV. Recommendation Pipeline

When a student submits their assessment, the backend serializes the skill and interest data into a JSON payload and forwards it to the Flask endpoint over HTTP. The Python service deserializes the payload, builds the user's feature vector, computes cosine similarity against all 85 career vectors, and runs a k-NN lookup with k set to 10. The top five results are returned with their similarity scores converted to percentages.

The backend receives these results, stores them in the database with a timestamp, and returns them to the frontend. Caching the results serves two purposes: it reduces latency on the recommendations page (which can simply fetch the stored result rather than re-running the model) and it creates a historical record that allows students to see how their recommendations evolve over time as they update their skills.

On the frontend, each recommendation is rendered as a card showing the career title, the match percentage displayed as a visual progress bar, the key matching skills highlighted in a tag cluster, a brief career description, and a link to further reading. This presentation was deliberately designed to be transparent — students can see exactly why a career was recommended rather than receiving an opaque list.

V. System Workflow

The end-to-end workflow proceeds through six steps. First, the student registers and creates a profile. Second, the backend validates credentials and issues an authentication token. Third, the student completes the skill and interest assessment. Fourth, the submitted data is persisted to the database. Fifth, the AI module processes the profile and generates ranked career suggestions. Sixth, the results are displayed on the

frontend recommendations page with match scores and supporting detail.

V. EVALUATION AND RESULTS

I. Pilot Study Setup

The system was evaluated through a controlled pilot study conducted with 47 undergraduate students from the Department of Information Technology at BIET, Lucknow. Participants were in their third or final year of their B.Tech program, an age group actively thinking about career options. Each participant used the platform independently on a lab computer, completing registration, the skill assessment, and reviewing their recommendations in a single session lasting approximately 25 minutes.

Following the session, participants completed two instruments: a custom relevance questionnaire and the System Usability Scale (SUS). The relevance questionnaire asked whether the top recommendation matched their own sense of their strengths (5-point scale), whether the recommended careers were ones they had considered before (yes/no), and whether they felt they had learned something new about career options (yes/no). The SUS, a validated 10-item questionnaire, provided a standardized usability score.

II. Quantitative Findings

Eighty-three percent of participants rated the top career recommendation as relevant or highly relevant to their skill profile. This is a strong result for a dataset-driven system without access to behavioral data or historical usage patterns. System Usability Scale scores averaged 76.4 out of 100, placing the platform in the 'good' range of the standard SUS benchmark scale [12]. Scores above 70 are generally considered acceptable for real-world deployment, and scores above 85 are considered excellent.

Sixty-one percent of participants reported that at least one of their top-five recommendations was a career path they had not previously considered, suggesting that the system is surfacing genuinely new information rather than simply confirming what students already know. This matters because simply validating existing beliefs — while comfortable for the user — provides limited value as a guidance tool.

A comparison against a random-ordering baseline showed statistically significant improvement in relevance ratings ($p < 0.01$, two-tailed paired t-test, $n = 47$). While the sample is not large enough to make strong generalizations, it provides sufficient evidence that the recommendation logic is doing meaningful work beyond chance.

Table 2. Summary of pilot study evaluation results.

Metric	Value
Total participants	47
Recommendation relevance (top-1)	83%
SUS score (avg.)	76.4 / 100
New career discovered	61%
Significance vs. baseline	$p < 0.01$
Avg. session duration	~25 minutes

III. Qualitative Observations

Beyond the numbers, several observations emerged from watching students interact with the system. Many participants spent more time on the results page than on the assessment itself, scrolling through the career cards and clicking on skill tags to understand why a particular career was recommended. This suggests that the transparency of the recommendation — showing matching skills explicitly — encourages deeper engagement with the output.

A recurring piece of feedback was that the assessment felt longer than expected. Students who had many skills to report spent more time on the form than the 25-minute average, occasionally skipping sections to finish. This points to a user experience issue in the assessment design that will be addressed in the next version through progressive disclosure and smarter default selections.

VI. LIMITATIONS AND ETHICAL CONSIDERATIONS

I. System Limitations

The most significant limitation of the current system is its dependence on accurate self-reporting. The quality of a recommendation is directly proportional to the quality of the input: students who underestimate their skills receive narrower recommendations, while students who overestimate them may receive suggestions that are not realistic. Addressing this requires either behavioral validation (observing what the student can actually do) or triangulation with external data sources such as academic transcripts or project portfolios — neither of which is currently implemented.

The career dataset, while carefully curated, reflects a snapshot of the job market. Career relevance shifts as industries evolve: roles that are prominent today may be automated or reshaped within a few years. The dataset requires periodic review and updating, which places an ongoing maintenance burden on the institution deploying

the system. Without this upkeep, the recommendations will gradually drift out of alignment with real market conditions.

Finally, the pilot sample was drawn exclusively from IT students at a single institution. This limits the external validity of the quantitative results. Students from different disciplines, institutions, or geographic contexts may interact with the system very differently.

II. Ethical Considerations

Any system that influences a person's career decisions carries ethical weight. The AI-Powered Career Mentor was designed with several safeguards in mind. First, all recommendations are presented as suggestions, not directives. The platform explicitly tells users that the match score reflects skill similarity, not a prediction of success or happiness in that career. Second, the recommendation logic is explainable: students can see which of their skills contributed to each match, which allows them to question and discuss the output rather than accepting it passively.

Data privacy is handled through standard web security practices: password hashing, token-based authentication, and HTTPS transport. User assessment data is stored only for the purpose of generating and displaying recommendations and is not shared with third parties. A future version of the system will incorporate a formal privacy policy and explicit user consent workflows in line with the Information Technology (Amendment) Act 2008 and evolving data protection norms in India [13].

Algorithmic fairness is a concern that warrants further study. The career dataset was not audited for demographic bias, and it is possible that the skill profiles encoded in the dataset reflect historical patterns that could disadvantage certain student groups. This is an area where explainable AI methods and adversarial fairness testing should be applied in future iterations.

VII. FUTURE WORK

The current system establishes a functional baseline that can be meaningfully extended in several directions. The highest-priority enhancement is the integration of NLP-based resume parsing. Rather than requiring students to manually check skill boxes, the system would extract skills from an uploaded resume using named entity recognition and skill classification models, pre-populating the assessment form and reducing friction significantly. This would also allow the system to capture nuanced skill descriptions that a checkbox form cannot accommodate.

The second planned enhancement is real-time skill gap detection. Once a student's target career is identified,

the system would compute the difference between their current skill vector and the career's required skill vector and present the missing skills as a prioritized learning roadmap, complete with links to relevant online courses and certifications. This transforms the platform from a recommendation tool into an ongoing development companion.

Third, the recommendation engine will be upgraded from k-NN with cosine similarity to a hybrid model combining content-based and collaborative filtering signals. As the user base grows, co-occurrence patterns in assessment data will provide collaborative signals that can improve recommendations for students whose profiles fall in underrepresented areas of the skill space. Over time, a larger dataset will also enable training of a lightweight neural recommendation model.

Fourth, an administrative dashboard will be developed for institutional use. Faculty and placement officers will be able to view anonymized, aggregated skill distribution data across cohorts, identify which career domains students are most aligned with, and plan curriculum adjustments accordingly. This closes the feedback loop between the guidance platform and the institution's educational strategy.

Fifth, accessibility and multilingual support will be extended. A Hindi language interface is planned for the next release, which will make the platform usable by students more comfortable in their regional language — an important consideration in the context of Uttar Pradesh, where the majority of engineering colleges serve students from Hindi-medium educational backgrounds.

VIII. CONCLUSION

This paper presented AI-Powered Career Mentor, a full-stack platform designed to bring data-driven career guidance to undergraduate students in academic institutions that lack the resources for large-scale human counseling. The motivation for this work came directly from the scale of the problem: millions of students graduate from Indian engineering colleges each year into a job market that increasingly demands not just technical competence but career clarity and direction.

The system addresses this by combining a well-structured skill assessment, a curated 85-profile career dataset, and a transparent recommendation engine that shows students exactly why each career was suggested. A pilot study at BIET, Lucknow, demonstrated that the approach is practically viable: participants found the recommendations relevant, the system usable, and the experience informative. Over 60% discovered career options they had not previously considered.

What makes this work more than a prototype is its modular, maintainable architecture. Each component —

frontend, backend, database, and AI module — can be upgraded independently. The dataset can be extended as new careers emerge. The recommendation engine can be replaced with a more powerful model as data accumulates. This extensibility is what gives the platform a realistic path from pilot to production.

The problem of career uncertainty among Indian graduates is too large and too important to be solved by any single tool. But platforms like AI-Powered Career Mentor can be one part of the answer — not replacing the human advisor, but making thoughtful, personalized guidance available to every student who needs it, regardless of which institution they attend or which city they come from.

ACKNOWLEDGEMENTS

The authors sincerely thank Prof. Gaurav Srivastava, Department of Information Technology, BIET, Lucknow, for his patient and thorough guidance throughout this project. His feedback at every stage — from the initial system design to the final evaluation — shaped this work in ways that are difficult to overstate. The authors also thank the student volunteers of BIET who gave their time to participate in the pilot study and provided honest, constructive feedback on the platform.

REFERENCES

- [1] Mercer-Mettl, "India Graduate Skill Index 2025," Mercer, New Delhi, India, 2025.
- [2] Unstop, "Unstop Talent Report 2025: Engineering Graduate Employment Outcomes," Unstop, New Delhi, India, 2025.
- [3] Wheebox, "India Skills Report 2024: Impact of AI on the Future of Work, Skilling and Mobility," Wheebox, Gurugram, India, 2024.
- [4] J. Doe, A. Smith, and B. Johnson, "A survey on AI-based career recommendation systems," *J. Artif. Intell. Res.*, vol. 45, no. 2, pp. 123-145, 2023.
- [5] S. Parsons, "The future of career guidance: Integrating technology with human-centered counseling," *J. Career Dev.*, vol. 48, no. 3, pp. 220-234, 2021.
- [6] ScienceDirect, "Representing jobs and job seekers in AI-based recommender systems: A literature review," *Expert Syst. Appl.*, 2026. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0952197626007311>
- [7] J. Li, X. Wang, and Y. Zhang, "Matrix factorization for student-career matching using alumni outcome data," in *Proc. ACM Conf. on Learning at Scale*, 2020, pp. 87-96.
- [8] P. Nithyanandam and T. Ramesh, "Content-based career recommendation using skill taxonomy vectorization," *Int. J. Comput. Appl.*, vol. 176, no. 12, pp. 34-41, 2020.
- [9] S. Soni and S. Shah, "Job recommendation systems through AI and machine learning," *Int. J. Basic Appl. Sci.*, vol. 14, no. 3, 2025.

- [10] L. Zhang, M. Wang, and Y. Liu, "Deep learning approaches for personalized career path prediction," IEEE Trans. Neural Netw. Learn. Syst., vol. 34, no. 5, pp. 987-999, 2024.
- [11] J. Thomas, "User-centric design in career recommendation systems," J. Hum.-Comput. Interact., vol. 39, no. 3, pp. 150-160, 2022.
- [12] J. Brooke, "SUS: A quick and dirty usability scale," in Usability Evaluation in Industry, P. W. Jordan et al., Eds. London, UK: Taylor & Francis, 1996, pp. 189-194.
- [13] N. Patel and M. Shah, "Data privacy challenges in AI career counseling platforms," J. Inf. Secur., vol. 18, no. 1, pp. 30-40, 2023.

ABOUT THE AUTHORS

VidyaSagar is a final-year undergraduate student in the Department of Information Technology at Bansal Institute of Engineering and Technology (BIET), Lucknow. His research interests span machine learning, full-stack web development, and the application of AI in education and career planning. He conceptualized, designed, and led the development of the AI-Powered Career Mentor platform as his undergraduate capstone project, under the supervision of Prof. Gaurav Srivastava. He can be reached at vidyasagararya41@gmail.com.

Prof. Gaurav Srivastava is a professor in the Department of Information Technology at BIET, Lucknow. His research interests include artificial intelligence, software engineering, and educational technology. He has guided multiple undergraduate and postgraduate projects in the domain of intelligent systems and data-driven applications, and brings several years of experience in academic mentorship in the field of computer science and information technology.