2026 Volume: 02 Issue: 01 www.ijrpa.com ISSN 2456-9995 Review Article

International Journal Research Publication Analysis

Page: 01-10

# TEAM SANGAM AI: SMART HACKATHON TEAM FORMATION SYSTEM

***Aman Kumar, Anamika Raj, Md. Alim**

Department of Computer Science and Business System Engineering Oriental Institute of Science and Technology, Bhopal, India.

**Article Received: 05 December 2025**
**Article Revised: 25 December 2025**
**Published on: 13 January 2026**

***Corresponding Author: Aman Kumar**
Department of Computer Science and Business System Engineering Oriental Institute of Science and Technology, Bhopal, India.

DOI: https://doi-doi.org/101555/ijrpa.5496

## ABSTRACT

Hackathon success critically depends on balanced, interdisciplinary teams. Traditional methods (self-selection or random grouping) often produce skill imbalances and poor collaboration. **Team Sangam AI** addresses this by using AI to form teams based on participants' technical skills, interests, and soft-skill profiles. Resumes and profiles are parsed via NLP to extract expertise, then participants are clustered (e.g., K-Means) into heterogeneous teams covering complementary skills. Optionally, an LLM (such as the OpenAI API) analyzes text responses or psychometric inputs to infer personality traits, ensuring interpersonal compatibility. Early results suggest that this data-driven approach yields more equitable, well-rounded teams and improves outcomes (since balanced teams outperform random onesjournals.mriindia.com). The platform also records historical team performance to continually refine its matching models. This paper presents the architecture of Team Sangam AI, a proposed tech stack, potential training datasets, and a step-by-step workflow. The system aims to enhance hackathon experiences by promoting diversity, fairness, and collaboration, thereby advancing team formation beyond static manual methods.

## INTRODUCTION

Hackathons are intensive, collaborative events where diverse skills must quickly coalesce into functioning teams. Yet, ad-hoc team formation (e.g. random grouping or friends teaming up) frequently yields imbalanced teams and sub-optimal performancejournals.mriindia.comijrti.org. Research shows that *balanced and diverse teams* – with complementary technical skills and interests – consistently produce higher creativity

and efficiency than randomly formed teamsjournals.mriindia.com. Similarly, consideration of soft skills (communication, personality, teamwork) has been shown to impact group harmony and successmdpi.comijrti.org. In practice, however, existing hackathon platforms largely ignore these factors, leading to unequal workload distribution and increased conflictijrti.orgijsrem.com.

To address these gaps, **Team Sangam AI** proposes an AI-driven team formation system. It profiles each participant (via resume/CV input or questionnaires), extracts both hard skills and indications of soft skills (through NLP analysis), and then algorithmically assembles teams. By leveraging clustering and machine learning, Team Sangam AI ensures that technical expertise is well distributed and personality traits are compatible. This approach contrasts with purely manual selection and static rules, providing a data-driven, dynamic solution that learns from past hackathons. The following sections review relevant literature on team formation algorithms, psychometric profiling, and skill matching techniques, identify key gaps, and then describe our proposed system architecture, datasets, and workflow.

## Literature Review & Gap Analysis

**Team Formation Algorithms:** Prior work spans various computational team formation methods. Systematic reviews show teams are typically formed by optimizing attributes like expertise, communication, or preferencesbusiness-and-management.orgfrontiersin.org. For example, graph-based and genetic algorithms have been used to match individuals to rolesbusiness-and-management.org, and clustering techniques (e.g. K-Means) have been applied to group students into heterogeneous project teamsijsrem.combusiness-and-management.org. In crowdsourcing contexts, most platforms rely on algorithms to automate team grouping based on user profiles and task requirementsfrontiersin.org, though purely top-down methods can alienate participantsfrontiersin.org. A notable hackathon case study contrasted open vs. algorithmic assignment: open assignment produced inconsistent skill mixes, while a naive algorithm led to dissatisfaction and reassignmentsdribdat-impetus.eu-central-1.linodeobjects.com. These findings underscore the need for smarter, inclusive algorithms in hackathons.

**Psychometric Profiling:** A growing body of research integrates personality into team formation. Studies using the Big Five model in student projects found that higher extraversion and conscientiousness (and lower neuroticism) correlate with better team performancemdpi.com. Predictive models (e.g. gradient boosting) have been developed to estimate team satisfaction from member personalitiesmdpi.com. However, most practical

systems do not leverage such insights. While personality tools (e.g. MBTI or CATME surveys) exist for classroom teams, hackathon platforms typically ignore these soft factorsijrti.orgarxiv.org. This gap suggests an opportunity: incorporating psychometric or language-based assessments into hackathon team matching could enhance cohesion.

**Automated Skill Matching:** In recruitment and education, AI-driven resume parsing and matching have proven effective. For instance, *Resume2Vec* uses transformer embeddings to align resumes with job descriptions, outperforming keyword-based systems by ~15%mdpi.com. The approach captures semantic context across skills (e.g. different wording for similar competencies)mdpi.com. Similarly, specialized hackathon platforms ("Hackomates", "TalentBridge") employ machine learning and NLP: they parse user profiles via tests or APIs (e.g. GitHub/LinkedIn), extract skill keywords, and match participants to projectsijrti.orgijsrem.com. Key methods include contextual NLP (e.g. KeyBERT) for skill extraction and K-Means clustering to group complementary profilesijsrem.com. Early implementations report more balanced teams and higher engagementijrti.orgijsrem.com.

**Gap Analysis:** Despite these advances, key gaps remain:

- **Neglect of Soft Skills:** Existing hackathon team builders focus on technical skills, neglecting personality or teamwork traitsijrti.orgarxiv.org. Without soft-skill analysis, team synergy can suffer.
- **Static, Manual Approaches:** Many organizers still form teams by sign-up lists or manual assignment, which is error-prone and doesn't scaleijsrem.comjournals.mriindia.com. Automated tools often lack adaptability (few systems update team logic based on feedback).
- **Imbalanced Skill Composition:** Studies note that imbalanced teams (e.g. clustering of similar skillsets) underperformjournals.mriindia.comijrti.org. Current systems do not guarantee optimal skill diversity in each team.
- **Lack of Learning from History:** Past team success patterns are rarely used. No integrated feedback loop exists to refine future matching based on which team compositions succeeded. Without learning, formation remains heuristic rather than evidence-based.

In summary, while team-formation algorithms and skill-matching tools existbusiness-and-management.orgijsrem.com, they often omit the human factors (personality/soft-skill) and

adaptability that hackathon contexts demandijrti.orgarxiv.org. Team Sangam AI aims to fill these gaps with a holistic, adaptive approach.

**Proposed Methodology (Solution)**

**System Architecture:** Team Sangam AI is designed as a modular AI-powered web platform. Users register and input their profiles (resumes, GitHub links, questionnaires). A **Resume/Skill Parser** (using NLP and keyword extraction) converts each profile into a feature vector of technical skills and interests. For example, resume text might be processed with transformer models (BERT/GPT) or NLP libraries (spaCy, KeyBERT) to identify skills and rolesijsrem.commdpi.com. Optionally, a **Soft-Skill Analyzer** uses a language model (e.g. via the OpenAI API) on free-text responses or survey answers to infer personality traits or teamwork preferences. These features, both hard and soft, feed into a **Team Clustering Engine**. We propose using **K-Means Clustering** to partition participants into *K* teams, optimizing for diversity: clusters maximize skill variance within each team while balancing experience levelsijsrem.comjournals.mriindia.com. After initial assignment, the system may allow a refinement phase (e.g. local swapping based on user preference). The entire platform is managed via a **Backend API** (FastAPI) with a document store (MongoDB) for flexible user and team data, and a **Frontend** (React.js/Next.js) for interactive dashboards. Regular retraining occurs: after each hackathon, team outcomes (e.g. project scores, peer reviews) are stored. A feedback loop retrains the clustering and weighting (or a separate ML model) so that future suggestions improve based on historical success.

| Component | Technology | Justification |
|---|---|---|
| Frontend | React.js / Next.js | Responsive UI/UX, server-side rendering, real-time interactivity for user inputs and team visualizations. |
| Backend | Python (FastAPI) | Flexible API layer for data processing, NLP integration, and serving model inferences. Python offers rich ML/NLP libraries. |
| AI Engine | Scikit-Learn (K-Means), OpenAI API | Scikit-Learn provides robust clustering and model tools; OpenAI (GPT) or similar LLM enables soft-skill/language analysis and resume embedding. |
| Database | MongoDB | Schema-flexible NoSQL store for user profiles, team records, and JSON NLP outputs; supports rapid iteration. |
| Authentication | Clerk or Firebase | Secure user management and authentication out-of-the-box, facilitating quick deployment. |

This stack ensures the platform can perform real-time processing (e.g. on signup), scale to many users, and easily incorporate AI/ML modules.

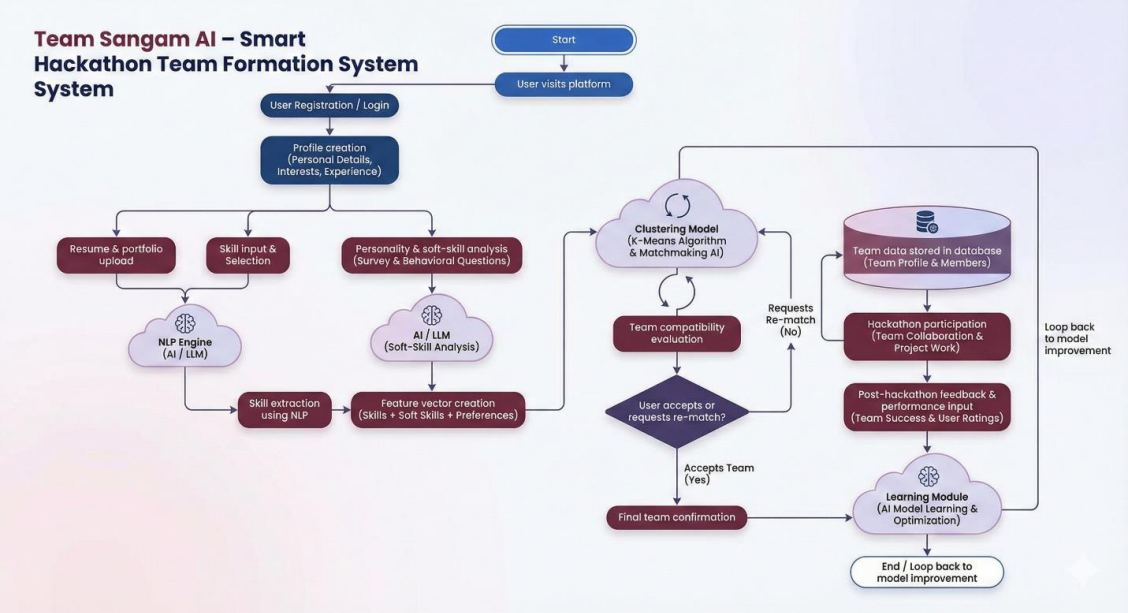
**Datasets**

To train and evaluate Team Sangam AI, we can leverage various public and synthetic datasets:

- **Hackathon Participant Data:** Kaggle hosts a *"Hackathon Participants"* dataset (200 high-school participants with skill/self-assessments) and *Devpost Hackathons* data (metadata on thousands of hackathon teams)kaggle.comkaggle.com. These can simulate real profiles and team outcomes.
- **GitHub Profiles:** Public datasets (e.g. Kaggle's GitHub User Analysis) provide developer attributes and project contributionskaggle.com. We can extract skill endorsements and activity metrics to mimic participant expertise.
- **LinkedIn/Resume Repositories:** While direct LinkedIn scraping is restricted, anonymized datasets or APIs (e.g. from Kaggle or academic sources) of professional backgrounds can inform skill extraction models. Synthetic data generation (with controlled skill/personality distributions) can supplement where privacy limits real data.
- **Developer Surveys:** Stack Overflow Developer Survey (2011–2024) includes respondents' self-reported skills, experience, and job satisfaction. These data provide realistic distributions of skills and interests.
- **Mock Hackathon Logs:** Historical data from past internal hackathons or online competitions (e.g. MIT's COVID-19 Challenge, university hackathons) – even if manually collected – can form a training pool. Synthetic teams (with known "success" labels) can be generated to bootstrap the clustering algorithm.

By combining such datasets, the AI models (cluster initialization, skill-weighting, and potentially a predictive model of team success) can be trained and validated before deployment. Importantly, the system remains adaptable: real hackathon results will continuously enrich the dataset.

## Workflow Description

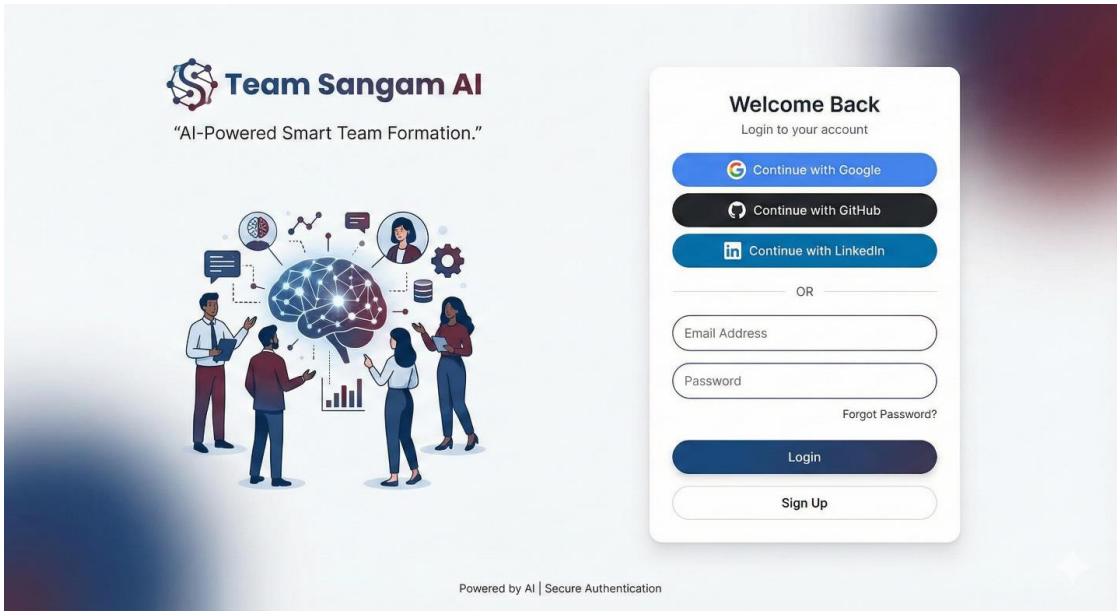



By iterating these steps, Team Sangam AI refines its matching logic. It adapts to the evolving community: as participants gain skills or as hackathon themes change, the system's model is continually retrained with fresh data.

**UI/UX Design Overview**

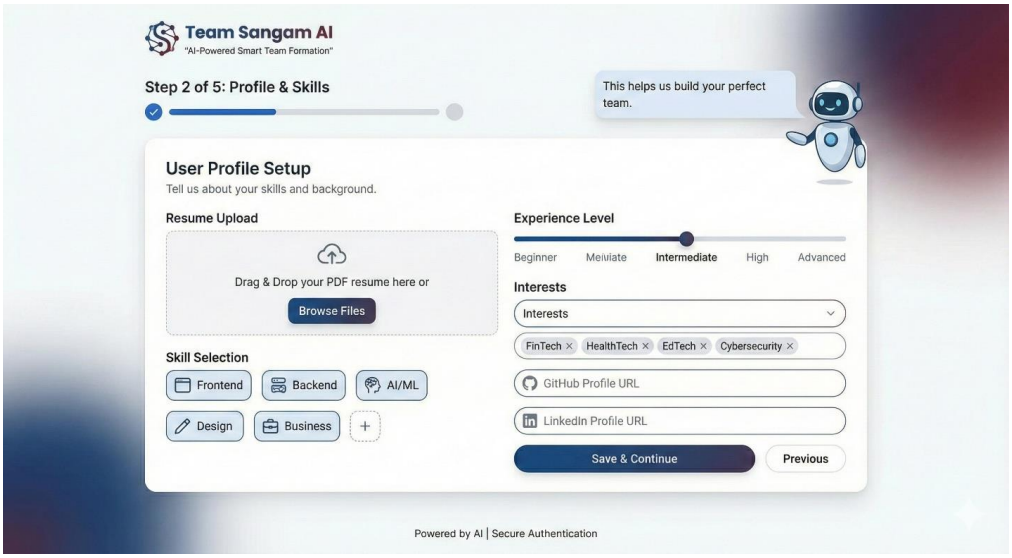
### 1. Login / Signup Page:

Provides secure user authentication through email or social login options to onboard participants.



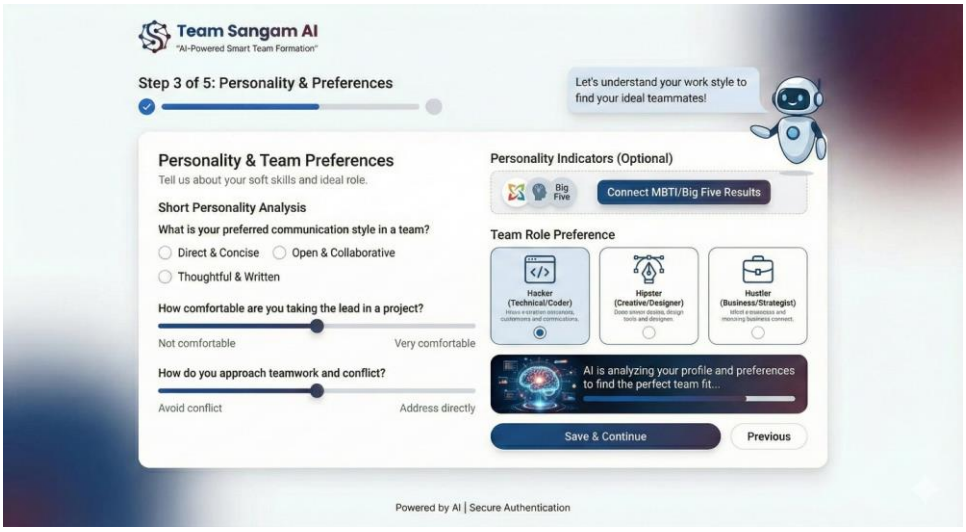
## 2. User Profile & Skill Input Page:

Collects users' technical skills, experience, interests, and portfolio details to build an accurate AI profile.



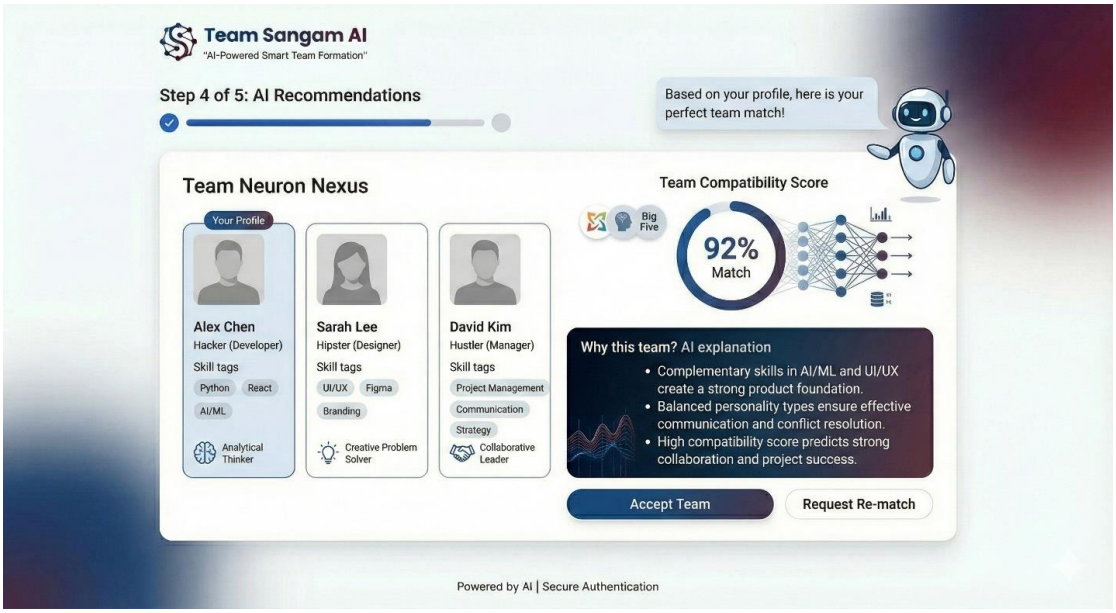

## 3. Personality & Team Preference Page:

Analyzes users' soft skills, communication style, and preferred team roles to improve team compatibility.



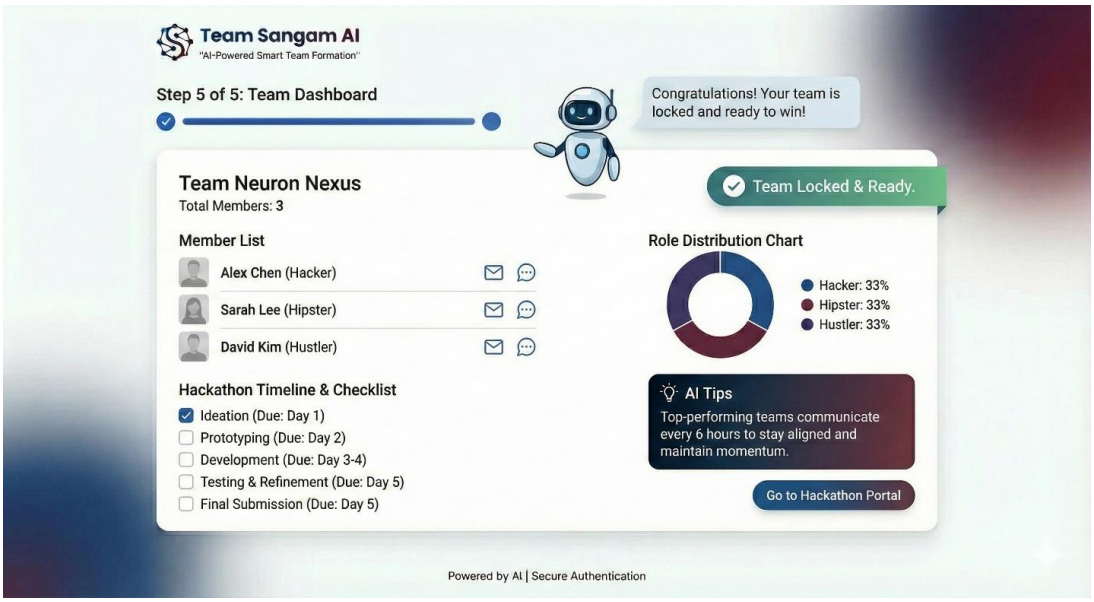

## 4. AI Team Recommendation Page:

Displays AI-generated, balanced team suggestions along with compatibility scores and matching explanations.




## 5. Final Team Dashboard Page:

Presents the confirmed team overview, role distribution, collaboration readiness, and hackathon status.




## Future Enhancements

- Integration of reinforcement learning to continuously improve team formation based on past hackathon outcomes.
- Addition of real-time collaboration analytics to monitor team interaction and provide adaptive AI suggestions.
- Mentor and expert matching module to assign domain experts to teams based on project themes.

- Integration with live hackathon platforms (Devpost, HackerEarth, GitHub) for seamless data synchronization.
- Implementation of privacy-preserving techniques such as federated learning for secure model training.
- Expansion to cross-institutional and global hackathons with scalable cloud-based deployment.
- Advanced explainable AI (XAI) features to improve transparency in team recommendation decisions.

## REFERENCES

1. D. Hevenstone, O. Lavrovsky, N. Endrissat, and O. Hümbelin, *"Team Formation and Inclusion in Hackathons: An Empirical Analysis and Research Proposal,"* Proc. of BaselHack 2021 (Workshop), 2021.
2. J. Vasiljević and D. Lavbič, *"A Data-Driven Approach to Team Formation in Software Engineering Based on Personality Traits," Electronics*, vol. 13, no. 1, p. 178, Jan. 2024.
3. M. Anusha, R. Nymisha, R. Madhumitha, V. Deepika, and A. Sowmya, *"Hackomates: Revolutionizing Team Formation Through Skill-Based Test Matching,"* Int. J. Res. Trends Innov., vol. 10, no. 4, Apr. 2025.
4. N. Gharate, I. Katariya, S. Kendre, S. Kothimbire, and N. Deshpande, *"Advanced Algorithms and Implementations for Talent Matching and Team Formation in Crowdsourced Engineering,"* Int. J. Sci. Res. Eng. Mgmt. Sci. (IJSREM), vol. 2, no. 7(S), pp. 98–107, Jul. 2025.
5. V. S. Nalawade, P. S. Patil, A. D. Gaikwad, and P. S. Panhalkar, *"Smart Team Builder for College Hackathons: A Comprehensive Survey,"* Int. J. Adv. Comput. Eng. Commun. Tech., vol. 14, no. 1, pp. 742–745, Nov. 2025.
6. R. V. K. Bevara *et al.*, *"Resume2Vec: Transforming Applicant Tracking Systems with Intelligent Resume Embeddings for Precise Candidate Matching," Electronics*, vol. 14, no. 4, p. 794, 2023.
7. G. Stavrou, V. Adamidis, S. Papathanasiou, and C. Tarabanis, *"Team Formation: A Systematic Literature Review,"* Int. J. Bus. Sci. Appl. Manag., vol. 18, no. 2, 2023.
8. G. Sujatha, B. T. Gajawada, M. Pallam, and S. S. Sai Sidhartha, *"JobFitAI: Automated Skill-Based Classification and Matching of Candidates to Relevant Job Openings,"* J. Sci. Eng. Tech. Mgmt. Sci., vol. 2, no. 7(S), pp. 98–107, Jul. 2025.

9. F. L. Vinella, J. Hu, I. Lykourentzou, and J. Masthoff, *"Crowdsourcing Team Formation with Worker-Centered Modeling,"* Frontiers in Artif. Intell., vol. 5, 2022.
10. M. Najafian *et al.*, *"Hackathons as an Educational Strategy for Shaping Soft Skills in Software Engineering,"* (Unpublished manuscript, 2024).