Optimizing Startup Team Composition: A Generative AI and Genetic Algorithm–Based Approach

Francesco Ferrati and Moreno Muffatto
University of Padova, Department of Industrial Engineering, Italy

francesco.ferrati@unipd.it
moreno.muffatto@unipd.it

Abstract: Founding team composition is widely acknowledged as a key determinant of startup ventures success. Entrepreneurship scholars have examined founding teams' characteristics through various perspectives, including human and social capital, psychological traits, personality, and diversity. Despite the recognized importance of team composition, traditional team formation methods often rely on convenience or heuristic-based combination. These approaches frequently overlook the systematic alignment of critical attributes such as complementary skill sets, personality compatibility, and shared motivations, leading to suboptimal outcomes and ultimately impacting team dynamics and performance. Given the transformative impact of generative Artificial Intelligence (AI) across multiple entrepreneurship domains, we investigate how Large Language Models (LLMs) can enhance the process of founding team formation by making the most of individual profiles. By employing advanced prompt engineering techniques, we properly instructed an LLM to develop a genetic algorithm to iteratively optimize team formation by evaluating and refining groups of three based on multiple criteria. The designed algorithm considers multiple sets of team configurations and efficiently converges on team assignments by applying selection, crossover, and mutation over subsequent generations to maximize demographic variety, hard and soft skills variety, personal value and motivation similarity and personality compatibility (based on the Big Five model). To facilitate reproducibility, we provide detailed prompt engineering strategies for implementation. To validate our method, we conduct an empirical study within a higher education entrepreneurship course where students are required to work in teams. We evaluated each team composition across seven metrics and assigned each team an overall score. Results suggests the effectiveness of our AI-assisted team formation approach. Our study addresses the growing interest in integrating generative AI into entrepreneurship research, offering insights to multiple stakeholders. While we tested our approach in the educational context, its applicability extends also to practitioners, including aspiring entrepreneurs searching for co-founders, startup studios forming founding teams, and organizations enhancing recruitment strategies.

Keywords: Startup, Team, Generative Artificial Intelligence, Large Language Model, Chatgpt, Genetic Algorithm

1. Introduction

When investigating the underlying reasons of startup failures, a flawed founding team frequently emerges as a key contributing factor (Wasserman, 2012). Nevertheless, many entrepreneurs still form teams primarily based on convenience rather than through a systematic combination of individuals' critical attributes (Ruef, Aldrich and Carter, 2003). Scholars has long debated whether team composition holds greater significance than other dimensions of nascent entrepreneurial ventures—such as the product or service, market potential, or financial structure—often considering these attributes as distinct, and framing the discussion through the so-called "jockey versus horse" controversy (MacMillan, Siegel and Narasimha, 1985). However, as Carpentier and Suret (2015: 819) state, "only experienced entrepreneurs are able to present sound projects," which suggests a relationship between entrepreneurial experience and the overall quality of the venture. While seasoned entrepreneurs may cultivate suitable co-founder selection ability, novice founders may lack this experience, making the team formation process particularly challenging. Nevertheless, some principles can inform the formation of a solid founding team (Arendt, 2024; Jin et al., 2017).

The small size of founding teams—typically comprising two to four individuals—makes the integration of non-overlapping expertise particularly important. When each co-founder assumes exclusive ownership over a distinct functional area without redundancy—such as product development, design, or sales—the team benefits from a clear division of responsibilities resulting in a broader spectrum of expertise and increased efficiency in decision making, learning curves, and coordination (Roberts, Negro and Swaminathan, 2013; Xiao et al., 2020). At the same time, to deal with the high level of uncertainty in early-stage ventures, co-founders need to share a set of personal motivations—whether related to social impact, lifestyle balance, or financial returns. Such alignment facilitates collective resilience and decreases the likelihood of conflict arising from divergent understandings of success (Murnieks, Klotz and Shepherd, 2020; Shane, Locke and Collins, 2003). Moreover, a common set of core personal values—such as integrity, helpfulness, humbleness—enables co-founders to approach complex dilemmas in a consistent way and contributes to the development of the emerging organizational culture (O'Neil, Ucbasaran and York, 2022). Given that building a new venture demands extended commitment and resilience, the compatibility of team members' personalities also plays a key role in both team effectiveness and long-term cohesion. When individuals exhibit aligned personality traits—such as conscientiousness, emotional stability, and openness to experience—they tend to engage in more effective communication, manage conflicts assertively, and maintain morale during inevitable setbacks (Freiberg and Matz, 2023; McCarthy et al., 2023). Finally, demographic variety—in terms of gender, ethnicity, age, or socioeconomic background—introduces a wider spectrum of life experiences, cultural knowledge, and social networks thereby expanding the team's cognitive capacity. However, especially in creative or innovative industries, diversity is positively related to creativity only within an inclusive environment, suggesting that without genuine integration of diverse perspectives into collaborative decision-making, diversity alone may not yield performance gains (Vedres and Vásárhelyi, 2023)

Despite the availability of such guiding principles, aspiring entrepreneurs frequently tend to select business partners based on perceived personal similarity (Kaiser and Müller, 2015) thereby limiting their selection to individuals within their immediate social circle. This approach occurs not only in the real-world entrepreneurial context but also in the educational environments. In the classroom, for instance, instructors who design group-based assignments often observe students choosing to collaborate with friends or people they already know (Gompers, Huang and Wang, 2017). Although prior familiarity may intend interpersonal compatibility, it does not necessarily ensure that the resulting teams possess the optimal combination of hard and soft skills, values and motivation necessary for success.

To enhance the team formation process, we offer a novel approach based on generative Artificial Intelligence (AI) and genetic algorithms. As general-purpose technologies, Large Language Models (LLMs) are showing increasing applications across multiple entrepreneurial domains, both in research (Ferrati, Kim and Muffatto, 2024) and practice (Fossen, McLemore and Sorgner, 2024). Most recent LLMs—such as ChatGPT o3 and o4-mini-high—incorporate reasoning-based modes that provide users with the ability to address challenging problems through a structured, step-by-step process (Yao et al., 2023; Wei et al., 2022). These models not only elaborate problem-solving strategies but also feature the capacity to autonomously generate and execute programming code, allowing both the development and direct implementation of solutions (Hou and Ji, 2025). As a response to the academic community's calls to explore how generative AI can advance entrepreneurship in both research and practice (Kusetogullari et al., 2025), in this study we demonstrate the potential of LLMs in the creation of founding teams.

To encourage experimentation with our approach, we offer two main contributions. First, we introduce a reproducible LLM-driven genetic algorithm for team formation, accompanied by comprehensive implementation guidelines and the tailored prompts we designed for use by educators and practitioners. Second, we empirically test the algorithm's effectiveness through an experimental study in the educational context, evaluating the results using key indicators such as variety, similarity and compatibility.

2. Methodology

To evaluate the effectiveness of our approach, we conducted an experiment in an educational environment, specifically within a master's-level entrepreneurship course where students worked in teams to develop innovative entrepreneurial projects. The student population, drawn from various courses of study in engineering (e.g., computer science engineering, electronic engineering, biomedical engineering), provided an ideal case for application. The experiment was conducted in three main phases: (1) profiling participants through an online form, (2) designing and implementing the team formation algorithm using generative AI, and (3) evaluating the method performance. Starting with the self-profiling of single individuals, our method automatically forms balanced teams based on well-established criteria in academic research.

2.1 Data Gathering

The first step in the team formation process requires each individual to create a self-profile. A total of 45 students took part in the study, all of whom provided informed consent for their anonymized data to be used for research purposes. Each participant was asked to fill a specially designed form to describe themselves according to seven key attributes: two demographic factors (gender and nationality), two related to their capabilities (hard skills and soft skills), and three concerning their psychographic characteristics (personal values, motivations and personality traits).

Our sample consists of 64 percent male and 36 percent female students, revealing a significant gender imbalance. While most of the participants are from Italy (60 percent), the remaining 40 percent come from a variety of European and non-European countries, giving the study a valuable extent of international and intercultural diversity. To map a comprehensive insight into individual capabilities, we asked each participant to self-report five hard skills and five soft skills. Participants declared 156 unique hard skills, reflecting a wide range of technical and discipline-specific expertise across multiple domains. Similarly, they reported 116 distinct soft skills, covering a broad spectrum of interpersonal, cognitive, and emotional abilities. To capture the psychological drivers that influence individual behaviour and decision-making, we asked participants to list five core personal values and five primary sources of motivation, alongside completing a personality assessment. In total, we collected 91 personal values, representing a large variety of beliefs and priorities, and 135 unique personal motivations, corresponding to different drivers that inform participants' actions. For personality profiling, participants took an online test based on the Big Five model (McCrae and Costa, 1987) and reported their score for each of the five traits—openness, conscientiousness, extraversion, agreeableness, and neuroticism.

Table 1 presents an example of an individual profile format submitted by participants. To ensure consistency and facilitate the subsequent algorithmic processing, we instructed participants to enter their hard skills, soft skills, personal values, and personal motivations into the form using the specific syntax ['item 1', 'item 2', 'item 3', 'item 4', 'item 5'].

Table 1: Sample of an individual's self-profiling entry

Attribute	Example
Gender	Female
Nationality	Italian
Hard skills	['deep learning', 'python', 'full stack development', 'project management', 'user experience']
Soft skills	['problem-solving', 'critical thinking', 'emotional intelligence', 'communication skills', 'empathy']
Personal values	['integrity', 'kindness', 'curiosity', 'helpfulness', 'humbleness']
Personal motivations	['joy', 'growth', 'impact', 'challenges', 'financial reward']
Personality trait (Big Five)	Openness: 90, Conscientiousness: 72, Extraversion: 91, Agreeableness: 97, Neuroticism: 60

2.2 Designing and Generating the Algorithm

To build effective teams based on individual profiles, we structured the logic of our team-formation algorithm around three interdependent phases as outlined in Figure 1. First, we systematically classified the information provided by participants. Second, we defined and assigned archetypes based on these classifications. Third, we generated and applied a genetic algorithm to optimize team composition. For implementation, we employed ChatGPT using the o4-mini-high model, an advanced reasoning model particularly well-suited for programming-related tasks. To ensure the reproducibility of our process, we provide a detailed description of each phase, including the specific prompt engineering strategies we designed.

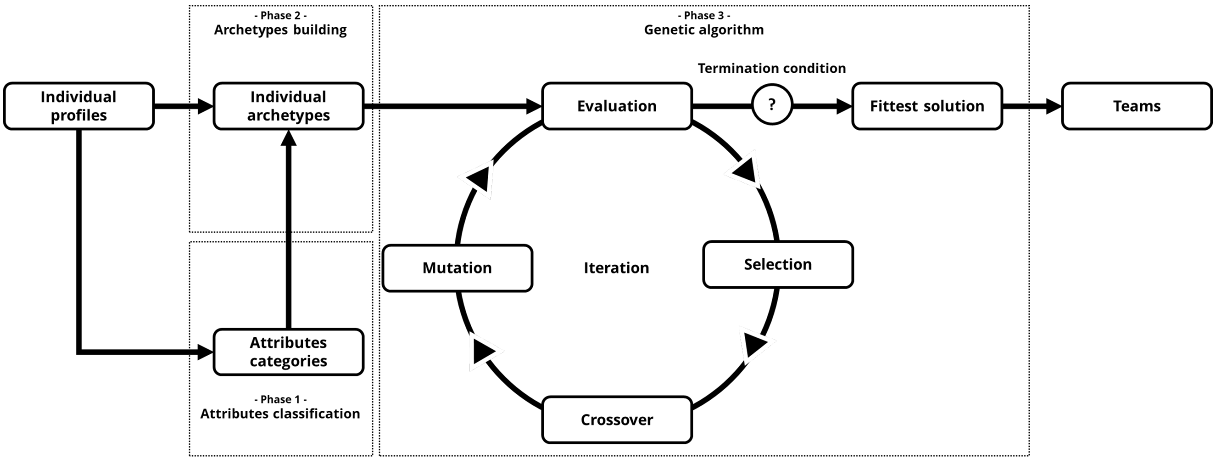

Figure 1: Operating logic of the team formation algorithm

2.2.2 Classifying Attributes

When collecting the data, we employed an open-ended format to gather detailed and authentic individuals' information, allowing participants to freely describe their profiles. This design intentionally produced considerable variability in responses. To systematically model each individual's key features, our algorithm first classifies these self-reported items into coherent thematic categories aligned with each variable. As shown in Table 2 this classification process was carried out by prompting ChatGPT according to the "role, context, task and instructions" prompt formula (Ferrati, Kim and Muffatto, 2024). We then double checked that each item was uniquely assigned to a single category to avoid duplication or omission.

At the end of this phase, 156 hard skills were classified into 22 categories, 116 soft skills into 8, 91 personal values into 9, and 135 personal motivations into 13. Figure 2 shows the top five categories of hard skills, soft skills, personal values, and personal motivations, ranked by the number of items in each.

Table 2: Prompt used to classify hard skills (to be adapted in the case of hard and soft skills, values and motivations)

Role: You are an expert data analyst specializing in hard skill taxonomy and classification. **Context**: You are provided with multiple lists containing various hard skills across different domains. **Task**: Analyse all the items in the provided skill lists and categorize them into clear, meaningful main categories. **Instructions**: • Group together skills that are closely related in meaning and domain. • Each skill must belong to one and only one category. • Do not use a "Miscellaneous" or "Other" category. • Ensure the categorization is highly precise and domain-specific. • Present the final output in JSON format, where each main category is a key and its associated skills are in an array. Here are the lists: [paste the lists here]

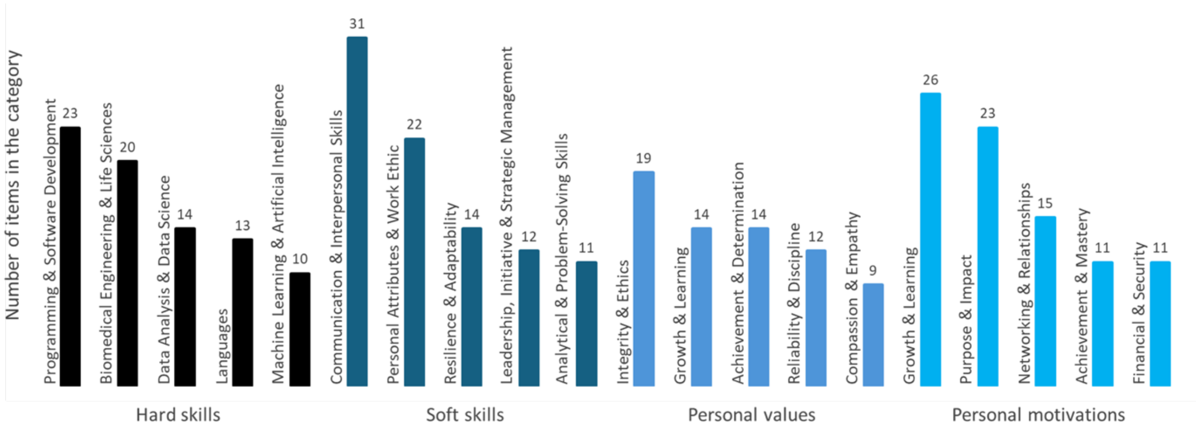

Figure 2: Top five categories of hard skills, soft skills, personal values, and personal motivations by number of items

2.2.3 Assigning Individual Archetypes

After mapping each self-reported items to its corresponding category, we employed ChatGPT to design and implement a strategy that assigns each participant a single representative value for each of the four dimensions: hard skill, soft skill, personal value, and motivation. By analyzing each individual's five submissions per dimension and selecting the most representative category, the algorithm transformed each dimension into a single archetype, significantly reducing profile complexity and improving computational efficiency in the subsequent team formation process. Table 3 shows the prompt we designed to execute this operation.

The algorithm generated by ChatGPT determines a primary category by applying a majority-vote mechanism. For example, in the case of an individual whose hard skills include ['Deep Learning', 'Computer Vision', 'Math', 'Programming', 'Photography']—then categorized as ['Machine Learning & Artificial Intelligence', 'Machine

Learning & Artificial Intelligence', 'Mathematics & Logic', 'Programming & Software Development', 'Design & Creative']—the algorithm assigned 'Machine Learning & Artificial Intelligence' as the hard-skill archetype since it appears most frequently.

Table 3: Prompt used to assign archetypes to variables

Role: You are an expert data analyst with experience in structured data transformation.

Context: You are working with a dataset of entrepreneur profiles, which includes the following columns: Hard skills, Soft skills, Personal values, and Personal motivations. Additionally, you have four separate JSON files, each mapping items from one of these columns to their corresponding categories.

Task: Enrich the dataset by categorizing profile attributes using the provided mappings and organizing the results for clarity and transparency.

Instructions:

- For each of the four columns, create a new column named categorized_[column_name] that assigns the most relevant category for each individual, based on the items in their list.
- Use a smart categorization strategy (e.g., based on frequency, priority, or semantic relevance) to determine the most appropriate single category per column when multiple items are listed.
- For transparency, add five intermediate columns per original column (e.g., hard_skill_cat_1, hard_skill_cat_2, ..., hard_skill_cat_5) to show the individual category assigned to each listed item.
- If a person has fewer than five items in a column, leave the remaining intermediate columns blank.
- Preserve all existing data and ensure the final output is clean, consistent, and well-organized.
- Once complete, return the updated dataset and provide it in Excel format (.xlsx) for download.

2.2.4 Implementing the Genetic Algorithm

After assigning archetypes to each participant's hard skills, soft skills, values, and motivations, we generated and implemented the team assignment algorithm using prompt-based techniques. Team formation can be approached algorithmically through several strategies, including random allocation, stratified balancing across multiple dimensions, or by applying specific constraints. Given the complexity of this optimization problem, we employed a genetic algorithm (GA) approach. Table 4 presents the prompt we designed to generate and execute the GA for team formation. GAs are inspired by the principles of evolutionary biology, where individuals with higher fitness are more likely to contribute to the next generation. The code generated by ChatGPT reveals the implementation strategy it applied.

Using GAs' own terminology, in our case the population is a set of fifty candidate solutions per generation. Each member of this population is an individual, and represents a complete ordering of all forty-five participants, which implicitly defines fifteen teams of three. The genetic information carried by an individual refers to a sequence of forty-five integer identifiers, each corresponding to a distinct co-founder profile. Reproduction and genetic recombination are performed by manipulating these identifier sequences.

A fitness function evaluates each individual based on seven criteria and assign them a fitness score, as outlined in Table 5. For gender, nationality, hard skills and soft skills, the algorithm aims to maximize variety; for values and motivations, the objective is to maximize similarity; for personality the goal is to maximize compatibility. Hard skills, soft skills, values and motivations are operationalized through the archetypes assigned in the prior phase. Personality traits, measured through Big Five scores, are treated as points in a five-dimensional space. To evaluate personality similarity, the algorithm computes the Euclidean distance between personality vectors and applies the transformation "1 - (average distance / maximum distance)", where "maximum distance" is the largest difference in personality that could exist between two participants. This strategy favours minimal divergence and maximizes the similarity between personalities.

To simulate evolutionary processes, GAs employ three core operators–selection, crossover, and mutation. Selection uses a tournament strategy: three individuals are randomly sampled from the current population, their fitness scores are computed and compared, and the fittest is selected as a parent to propagate its genetic information. Repeating this process fifty times yields the parent pool for the next generation. By mimicking reproduction, crossover combines genetic material from two of these parents to generate offspring by copying a segment from parent A's forty-five-element sequence and filling in the remaining positions in the child with the order-preserved elements from parent B. This order-preserving recombination enables the inheritance of partial team structures from one parent while ensuring genetic mixing by integrating elements of the other. Finally, mutation is introduced by randomly swapping two positions in the offspring's sequence with a 20%

probability, maintaining genetic diversity within the population and preventing from converging too quickly on suboptimal solutions

The algorithm repeats the evolutionary cycle—evaluation, selection, crossover, and mutation—for one hundred generations, then terminates the process and select the highest-fitness individual encountered throughout the run as the optimal assignment of the forty-five participants into fifteen teams.

Table 4: Prompt used to generate e genetic algorithm for team formation

Role: You are an expert in genetic algorithms and entrepreneurial team composition and dynamics.

Context: You are provided with a dataset containing detailed profiles of startup founders. The objective is to strategically form founding teams of 3 individuals each, based on factors proven in academic literature to influence team effectiveness in entrepreneurship.

Task: Design and execute a sophisticated genetic algorithm to optimally form 15 teams of 3 founders each, maximizing and balancing key team dynamics.

Instructions:

- Implement a full genetic algorithm pipeline including all essential phases: initialization, evaluation, selection, crossover, mutation, and convergence criteria.
- Optimize the following dimensions for each team: gender variety, nationality variety, hard skill variety, soft skill variety, value similarity, motivation similarity, personality compatibility (based on the Big Five personality traits).
- Ensure that your scoring and fitness functions reflect insights from academic literature on startup team effectiveness, especially regarding how to balance variety, similarity and compatibility.
- Once optimal teams are formed, append a new column to the dataset called assigned_team, indicating the team number (1–15) for each founder.
- Preserve all original data and return the updated dataset in Excel format (.xlsx) for download.

Table 5: Fitness Function applied by the genetic algorithm

Criterion	Score definition	Objective
Gender variety	1.0 if all 3 are different, 0.67 if 2, 0.33 if same gender	Higher: more diverse
Nationality variety	1.0 if all 3 are different, 0.67 if 2, 0.33 if same nationality	Higher: more diverse
Hard Skills variety	1.0 if all 3 are different, 0.67 if 2, 0.33 if same hard skills	Higher: more diverse
Soft Skills variety	1.0 if all 3 are different, 0.67 if 2, 0.33 if same soft skills	Higher: more diverse
Values similarity	Highest frequency of a shared value / 3	Higher: more similar
Motivations similarity	Highest frequency of a shared motivation / 3	Higher: more similar
Personality compatibility	1 - (average pairwise Euclidean distance between personalities / max distance)	Higher: more compatible

3. Results

To assess the effectiveness of our algorithmic approach we assessed the quality of each team using established diversity metrics (Harrison and Klein, 2007). Specifically, for each team, we calculated three distinct measures. First, we calculated Blau's index to assess variety in categorical variables such as gender, nationality, hard skills and soft skills. Second, for personal values and motivations, we employed the complement of Blau's index (1-Blau's index) to capture similarity. Third, to evaluate personality compatibility, we applied the formula [1 - (average pairwise Euclidean distance between personalities / max distance)], which measures the extent to which personality profiles within a team are aligned.

Table 6 reports the metric values for each team, alongside the theoretical minimum and maximum values derived from the properties of each indicator. To facilitate cross-team interpretation and indicators' comparison, we then normalized all indicators on a uniform scale from 0 to 1, as shown in Figure 3. When we examine the normalized values, we observe that the algorithm prioritized maximizing the variety of both hard skills (mean = 0.93; sd = 0.14) and soft skills (mean = 0.96; sd = 0.12) with perfect variety in 80% and 87% of cases, respectively. Similarity in personal values (mean = 0.60; sd = 0.34) and motivations (mean = 0.53; sd = 0.41) reaches the maximum level in 40% of the teams, while in three teams, the algorithm fails to ensure any shared motivation among members. Teams 3 and 7 are particularly noteworthy, as they exhibit perfect alignment in personal values despite showing zero similarity in motivations. This contrast illustrates how the GA may prioritize shared foundational beliefs over alignment in individual drivers when balancing criteria. Personality compatibility is relatively consistent (mean = 0.73; sd = 0.05) suggesting that the algorithm effectively aligned interpersonal dispositions. Considering demographic indicators, variety in gender (mean = 0.64; sd = 0.20) and nationality (mean = 0.78; sd = 0.16) appear limited in some teams due to demographic imbalances in the sample. In cases of low demographic variety, the algorithm still compensated by focusing on maximizing variety in skill domains and reinforcing similarity in personal values and motivations. These results reflect the trade-offs of founding team's composition. By integrating a multi-objective fitness function, our algorithm successfully addressed these tensions, producing practically balanced team configurations.

Table 6: Team metrics evaluation.

(a) Blau's index, (b) 1 – Blau's index, (c) 1 – (Average pairwise Euclidean distance/D_max)

	Gender variety	Nationality variety	Hard Skill variety	Soft Skill variety	Values similarity	Motivations similarity	Personality Compatibility
Metric	*(a)*	*(a)*	*(a)*	*(a)*	*(b)*	*(b)*	*(c)*
Effective K	*min(2,3)=2*	*min(12,3)=3*	*min(13,3)=3*	*min(7,3)=3*	*min(7,3)=3*	*min(8,3)=3*	–
Min	*0.000*	*0.000*	*0.000*	*0.000*	*0.333*	*0.333*	*0.000*
Max	*0.500*	*0.667*	*0.667*	*0.667*	*1.000*	*1.000*	*1.000*
Team 1	0.444	0.667	0.667	0.667	1.000	1.000	0.722
Team 2	0.444	0.444	0.667	0.444	0.556	0.556	0.729
Team 3	0.444	0.667	0.667	0.667	1.000	0.333	0.738
Team 4	0.444	0.444	0.667	0.667	1.000	0.556	0.733
Team 5	0.667	0.444	0.667	0.667	0.556	1.000	0.756
Team 6	0.444	0.444	0.444	0.667	0.556	1.000	0.753
Team 7	0.444	0.444	0.667	0.667	1.000	0.333	0.737
Team 8	0.444	0.667	0.444	0.667	0.556	0.556	0.689
Team 9	0.000	0.444	0.444	0.667	0.556	1.000	0.642
Team 10	0.444	0.444	0.667	0.667	0.556	1.000	0.809
Team 11	0.444	0.667	0.667	0.444	0.556	1.000	0.757
Team 12	0.444	0.444	0.667	0.667	0.556	0.333	0.619
Team 13	0.444	0.667	0.667	0.667	1.000	0.556	0.734
Team 14	0.444	0.444	0.667	0.667	1.000	0.556	0.662
Team 15	0.444	0.444	0.667	0.667	0.556	0.556	0.799

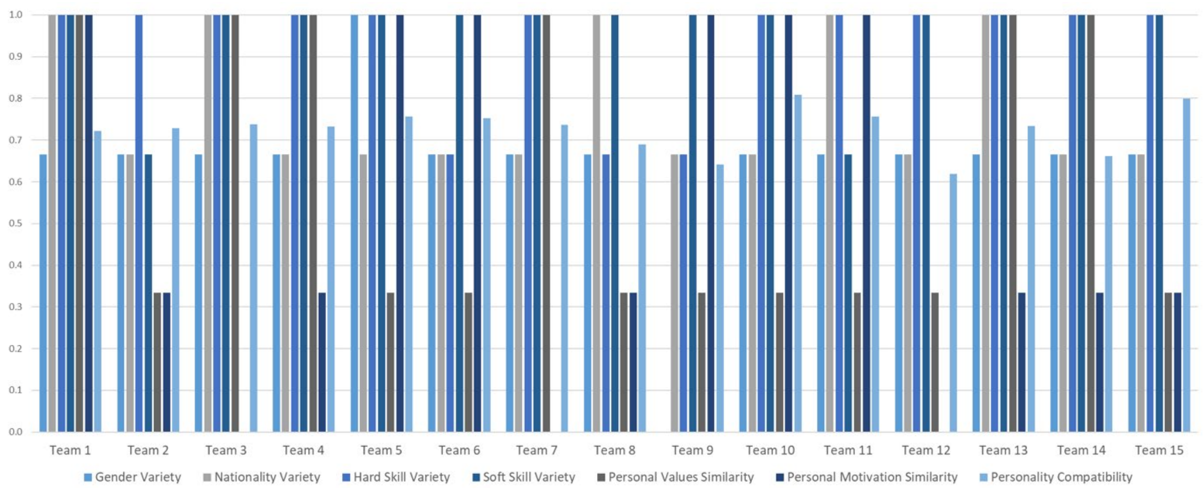

Figure 3: Team metrics evaluation, normalized to [0,1]

While the analysis of individual indicators allows for a disaggregated assessment of the algorithm's performance across distinct dimensions, evaluating the overall quality of each team requires a single, representative composite score. To this end, we employed the Weighted Root Mean Square (Weighted-RMS) Composite. This method involves squaring each metric—thereby amplifying the influence of larger deviations, both high and low—applying a weight to each squared value based on its relative importance, averaging these weighted squares by dividing by the sum of the weights, and finally taking the square root to return the score to the original metric scale. In our implementation, we distributed the weights to ensure a total contribution of 100%. Specifically, we assigned 25% each to hard-skill and soft-skill variety, 15% each to values and motivations homogeneity, 10% to personality compatibility, and 5% each to gender and nationality variety. Although this weighting scheme reflects our evaluation criteria, it can be adjusted based on alternative assessment frameworks.

As illustrated in Figure 4 the resulting Weighted-RMS composite scores for the fifteen teams range from 0.57 to 0.78 on a scale from 0 to 1. The median composite score is 0.68 (mean 0.67; sd = 0.06), indicating overall solid performance. In addition to the overall score, we visualize the weighted specific contribution of each team characteristic, highlighting how the mix of key elements leads to a balanced team configuration. To further evaluate the relative contribution of each individual factor to the overall score, Figure 5 expresses these contributions in percentage terms, thereby facilitating the identification of the most and least influential attributes associated with each team. Given the weights we assigned, we observe that, on average, the diversity of hard skills (mean 0.22; sd = 0.06) and soft skills (mean 0.23; sd = 0.06) exerts the strongest influence on the overall score. This is followed by the similarity of personal values (mean 0.19; sd = 0.10) and motivations (mean 0.18; sd = 0.12), then by personality compatibility (mean 0.12; sd = 0.02), with the least impact coming from diversity in nationality (mean 0.03; sd = 0.01) and gender (mean 0.02; sd = 0.01). However, when analysing the different teams individually, we note that their strengths can differ from the average trend. For example, the score of the best performing team (Team 1) is equally driven by the variety of hard and soft skills, each contributing 18 percent to the total score, but most significantly by the similarity of personal values and motivations, which together account for almost 50 percent. This case highlights that strong teams can emerge through synergistic mechanisms that emphasize different dimensions of team composition.

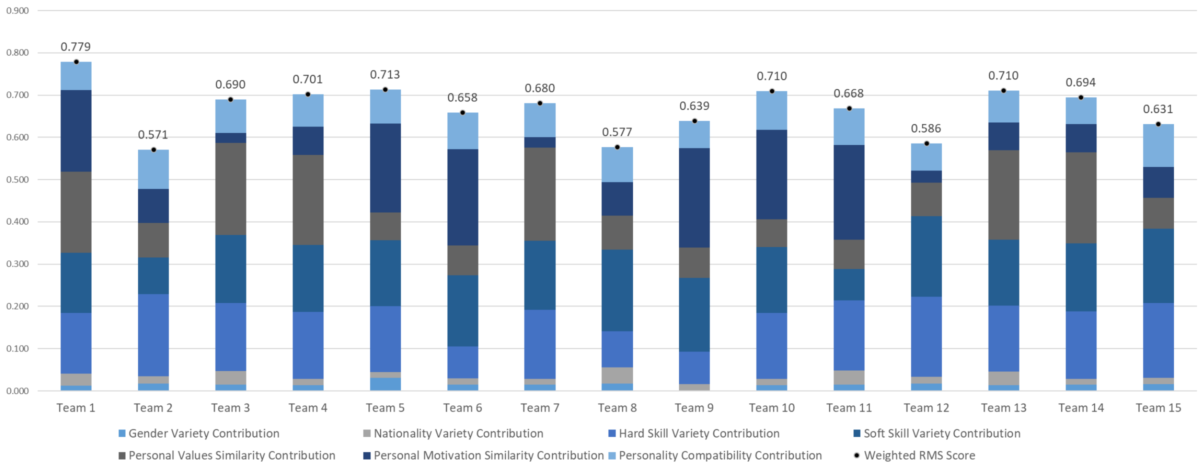

Figure 4: Weighted RMS Score, highlighting the weighted contribution of each attribute

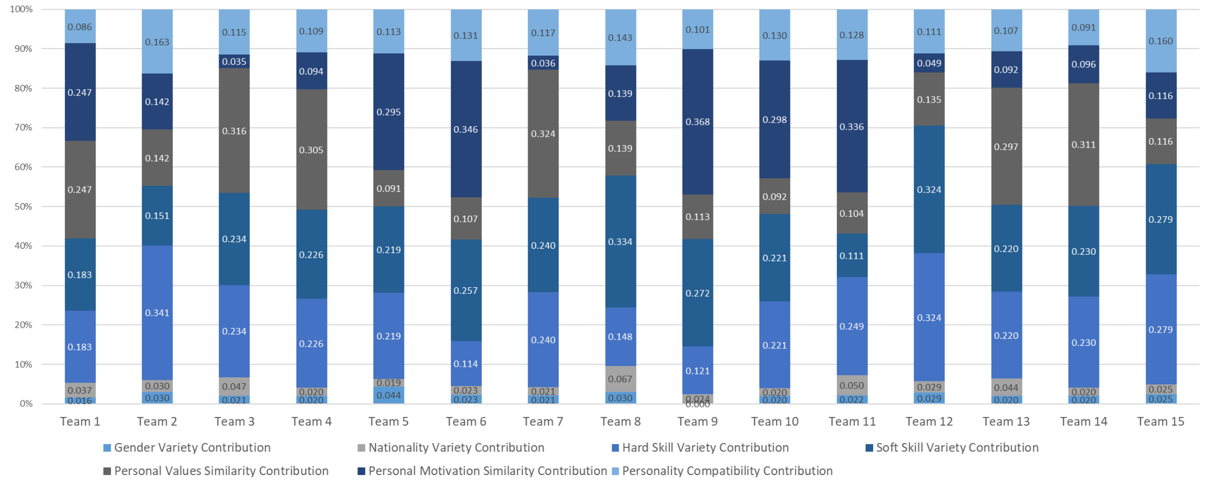

Figure 5: Weighted RMS Score, highlighting the weighted contribution of each attribute as a percentage of the score

4. Discussion and Conclusion

In this study we proposed a novel, replicable approach to entrepreneurial team formation that combines generative AI and genetic algorithms. Although prior research has acknowledged the importance of team composition, it has rarely offered systematic methods for building optimal founding teams. This work aims at contributing to filling this operational gap by demonstrating how algorithmic team formation can optimize key team characteristics simultaneously.

While we experimented with our approach within a higher education entrepreneurship course, it can also be transferred outside of the educational context. For aspiring entrepreneurs and startup studio managers, our findings indicate that generative AI can facilitate founder matching by efficiently analysing individual profiles and forming well-balanced co-founder teams. Similarly, human resource professionals may adopt this approach to enhance recruitment processes and optimize team composition within corporate settings. Moreover, while implementing the genetic algorithm, weighting parameters and selection criteria can be adjusted to meet specific pedagogical or professional objectives—for example, emphasizing creative heterogeneity in innovation-focused curricula or fostering value alignment in social impact projects.

Although being effective, our methodology also presents some limitations that offer opportunities for future development. For example, since we assessed the quality of the teams at the configuration stage rather than throughout the projects' lifecycles, future research should conduct controlled experiments to test whether AI-generated teams outperform teams formed organically by participants. Second, although our fitness function relied on a set of criteria informed by the literature, additional dimensions could be explored, as well as the influence of different weightings on team formation and performance. Third, since we relied on self-reported data, potential biases could have been introduced. To improve accuracy and mitigate social desirability effects, third-party assessments, peer evaluations, or behavioural data could be incorporated.

This work contributes to the ongoing discourse on entrepreneurial team formation by offering an actionable and customizable framework for building high-potential teams in both educational and professional environments.

Ethics Declaration

Given that the study relies on participants' self-profiling data, each individual provided informed consent for the use of their anonymized information for research purposes.

Declaration of Generative AI Usage

After completing the initial draft, the authors employed ChatGPT-4o and -o4-mini-high to refine the clarity and to implement the team formation algorithm as stated in the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for its contents.

References

Arendt, K. M. (2024). Team Formation Strategies Among Prospective Entrepreneurs–Evidence From a Large-Scale Survey Experiment. Entrepreneurship Education and Pedagogy, 7(1), 62-92.

Carpentier, C., & Suret, J. M. (2015). Angel group members' decision process and rejection criteria: A longitudinal analysis. Journal of Business Venturing, 30(6), 808-821.

Ferrati, F., Kim, P. H., & Muffatto, M. (2024). Generative AI in entrepreneurship research: Principles and practical guidance for intelligence augmentation. Foundations and Trends® in Entrepreneurship, 20(3), 245-383.

Fossen, F. M., McLemore, T., & Sorgner, A. (2024). Artificial intelligence and entrepreneurship. Foundations and Trends® in Entrepreneurship, 20(8), 781-904.

Freiberg, B., & Matz, S. C. (2023). Founder personality and entrepreneurial outcomes: A large-scale field study of technology startups. Proceedings of the National Academy of Sciences, 120(19), e2215829120.

Gompers, P. A., Huang, K., & Wang, S. Q. (2017). Homophily in entrepreneurial team formation (No. w23459). National Bureau of Economic Research.

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. Academy of management review, 32(4), 1199-1228.

Hou, W., & Ji, Z. (2025). Comparing large language models and human programmers for generating programming code. Advanced Science, 12(8), 2412279.

Jin, L., Madison, K., Kraiczy, N. D., Kellermanns, F. W., Crook, T. R., & Xi, J. (2017). Entrepreneurial team composition characteristics and new venture performance: A meta–analysis. Entrepreneurship theory and practice, 41(5), 743-771.

Kaiser, U., & Müller, B. (2015). Skill heterogeneity in startups and its development over time. Small Business Economics, 45, 787-804.

Kusetogullari, A., Kusetogullari, H., Andersson, M., & Gorschek, T. (2025). GenAI in Entrepreneurship: a systematic review of generative artificial intelligence in entrepreneurship research: current issues and future directions. arXiv preprint arXiv:2505.05523.

MacMillan, I. C., Siegel, R., & Narasimha, P. S. (1985). Criteria used by venture capitalists to evaluate new venture proposals. Journal of Business venturing, 1(1), 119-128.

McCarthy, P. X., Gong, X., Braesemann, F., Stephany, F., Rizoiu, M. A., & Kern, M. L. (2023). The impact of founder personalities on startup success. Scientific Reports, 13(1), 17200.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. Journal of personality and social psychology, 52(1), 81.

Murnieks, C. Y., Klotz, A. C., & Shepherd, D. A. (2020). Entrepreneurial motivation: A review of the literature and an agenda for future research. Journal of Organizational Behavior, 41(2), 115-143.

O'Neil, I., Ucbasaran, D., & York, J. G. (2022). The evolution of founder identity as an authenticity work process. Journal of business venturing, 37(1), 106031.

Roberts, P. W., Negro, G., & Swaminathan, A. (2013). Balancing the skill sets of founders: Implications for the quality of organizational outputs. Strategic Organization, 11(1), 35-55.

Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The structure of founding teams: Homophily, strong ties, and isolation among US entrepreneurs. American sociological review, 68(2), 195-222.

Shane, S., Locke, E. A., & Collins, C. J. (2003). Entrepreneurial motivation. Human resource management review, 13(2), 257-279.

Vedres, B., & Vásárhelyi, O. (2023). Inclusion unlocks the creative potential of gender diversity in teams. Scientific Reports, 13(1), 13757.

Wasserman, N. (2012). The founder’s dilemmas: Anticipating and avoiding the pitfalls that can sink a startup. Princeton University Press.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824-24837.

Xiao, Y., Dowejko, M. K., Au, K., & Hsu, A. J. (2020). “Jack-of-all-trades” with passion: Keener to pursue startup in a team?. Journal of Small Business Management, 58(4), 806-833.

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. Advances in neural information processing systems, 36, 11809-11822.