CareerVillage.org

Bellwether

Built for Learning: How Coach by CareerVillage Uses Artificial Intelligence for Career Development

By Marisa Mission, Michelle Croft, and Amy Chen Kulesa

JANUARY 2026

Series Overview: The integration of artificial intelligence (AI) into ed tech tools has raised myriad questions about how such advanced technology can both ease burdens for students and teachers and facilitate deep learning. Building on Bellwether's prior work examining how AI could amplify productive struggle and how to measure the impact of AI-powered ed tech tools, this case study series showcases those concepts in practice by spotlighting select organizations and describing their design approaches, trade-offs, and implementation choices. The case studies in this series are drawn from interviews conducted with organization leaders in summer 2025, and each profiled organization reviewed its case study for accuracy in October 2025. **Learn more by reading Bellwether's Built for Learning series.**

A Case Study on CareerVillage

Introduction

Guided by a mission to democratize access to career information and support, CareerVillage.org began as a community forum connecting teenage and postsecondary students to educators and working professionals.[1] After the release of the first generative artificial intelligence (GenAI) models in 2022, the organization saw an opportunity to use AI to scale students' access further. In 2023, CareerVillage began developing its own AI-powered career advising tool, and by 2024, the team released Coach, a text-based chatbot trained to assist students with career exploration and skill building. The product's goal is not just to answer students' questions but also to amplify the mindsets and behaviors that lead to successful careers. Chats are therefore focused on discrete career mentorship activities — such as exploring certain fields or roles, preparing for job interviews, or drafting résumés[2] — all of which CareerVillage built in partnership with experts and partner organizations.

Among the landscape of AI-powered ed tech tools, Coach by CareerVillage stands out, not only for its focus on career preparation rather than classroom learning but also for its:

1. Evidence-based impact model that includes short- and medium-term indicators to provide timely, useful data while waiting for longitudinal data collection.
2. Safety-first orientation throughout development, which establish strong guardrails to prevent model misbehavior.
3. Activity-specific rubrics developed by human experts that create benchmarks that allow for AI-scaled quality evaluations.
4. Right-sized amount of AI, optimized not just for cost and latency, but also for healthy teen engagement.

From Issue to Impact

As a domain that typically sits outside the K-12 core, career development is often overlooked — the time students can dedicate to it and the resources a district might spend on it are both limited by other priorities. In high school, where career development is most likely to be emphasized, counselors often struggle with unmanageable caseloads that do not allow for personalized attention or hands-on skill building.[3] These issues are even more severe for students from underserved backgrounds[4] and can be exacerbated when students lack career support or sources of information outside of school. Coach targets these gaps by 1) meeting learners where they are, outside the classroom; 2) helping learners practice concrete skills and behaviors necessary for career readiness; 3) tailoring advice and exploratory conversations to students' interests; and 4) leveraging CareerVillage's existing community forums to direct students to answers from real working professionals.

Coach focuses on building the mindsets and skills students need for the workforce.

CareerVillage created an impact model based on changing students' behaviors and mindsets, which can be measured directly and immediately. CareerVillage partnered with researchers to conduct a literature scan of six career readiness theories, through which they identified three measurable constructs of a career-ready mindset: interest, self-concept, and self-efficacy.[5] Any activity students do with Coach is mapped to one or more of those constructs, and when the chat is over, learners are asked a single yes-or-no question meant to measure whether the activity improved skills related to that construct.[6] Not only does this question help the CareerVillage team measure the effectiveness of the activity, but the data also create a short-term progress tracker for that student.

To track progress, students periodically complete the Career Adapt-Abilities Scale (CAAS), a 12-question survey that measures career "adaptability."[7] While this is not exactly the same as the three prior constructs, the team found evidence that all four concepts are linked; together, these measurements can provide a holistic picture of a student's career readiness.[8] Students complete the CAAS before engaging with Coach as well as at regular intervals throughout usage of the tool. This pre- and post-testing provides a more in-depth, standardized view of growth over time using a well-known survey instrument.

Throughout 2025, CareerVillage has also evolved its impact model to include medium-term outcomes such as certifications attained, internships secured, and job applications or offers — milestones that typically occur within three to six months of using Coach. To build the infrastructure necessary for collecting these metrics, CareerVillage has worked with its partners, which are typically national job training programs and institutions of higher education, on data-sharing agreements that connect user activity on Coach with those intermediate indicators. Using these data as leading indicators, CareerVillage can track and demonstrate Coach's impact in the near term as the team continues working toward tracking longer-term outcomes.

The organization's ultimate goal is to be able to collect the traditional longitudinal metrics of job placement, satisfaction, and earnings. However, these are outcomes that a student may not attain for six to 12 months after engaging with Coach, depending on their age, field of interest, or myriad other factors. This is a common challenge across the education and workforce development sectors: Even when organizations secure access to institutional data, that data may be inaccurate or outdated, and students who experience meaningful impact may not report back. To address these challenges, CareerVillage is leaning into collaborative efforts to connect disparate data. At the same time, the team is working to strengthen its internal research capacity to analyze data for partners, with the hopes of sharing its findings for others working in the space.

In the meantime, the near-term metrics for Coach are promising: A pilot program in August 2023 with nonprofit partner Generation.org found that Coach helped young adult users feel more optimistic about their job search.[9] More recently, surveys from June 2024 through January 2025 demonstrate that 85%-90% of

users saw improvement in their interest, self-concept, or self-efficacy.[10] By using an evidence-based model that directly links the use of technology to mindset shifts, CareerVillage has been able to measure Coach's impact beyond typical engagement outputs to demonstrate exactly how the AI-powered tool can boost student outcomes.

Using AI With Intention

Safety considerations shaped Coach's development from the beginning.

CareerVillage became a founding member of the EDSafe AI Alliance Industry Council,[11] which influenced its design choices and helped the team create guardrails for both the quality and safety of student interactions. CareerVillage Founder and Executive Director Jared Chung described the first version of those safety guardrails as entirely human-driven: Staff members read every AI-generated response to ensure chats stayed safe and were helpful for students.[12] Over time, the team developed standards and frameworks that allowed it to automate safety scans; these evolved into what it now calls "universal standards" that protect students' psychological safety across all chats.[13]

As Coach usage has grown, the CareerVillage team has added a layer of real-time safety monitoring using an AI model trained on the universal standards. A human moderation team still reviews flagged conversations and intervenes when necessary,[14] but Chung said that between the universal standards and the inherent guardrails built into newer models, it is "extraordinarily rare" for Coach to not meet safety guidelines.[15] If it happens, the team has found that "usually it's because [a user] is trying to make Coach say something inappropriate."[16]

Behind the scenes, Coach has several layers of benchmarking and evaluations to ensure students are getting high-quality responses.

Although students may experience a free-flowing, natural-feeling conversation, CareerVillage meticulously planned Coach activities to ensure that students receive high-quality support. Before releasing any activity, CareerVillage works with career counseling experts to create an activity-specific rubric with criteria grounded in coaching and learning best practices.[17] These rubrics then serve as benchmarks for high-quality responses. Outside of the activity-specific rubrics, there are additional criteria that measure coaching best practices across all chats; altogether, Chung noted there are more than 150 individual measures of conversation quality, and any conversation could be evaluated for multiple of those measures.[18]

To assess Coach's performance, CareerVillage uses a multilayered approach. The most basic level includes quick checks such as monitoring whether users thank Coach for its help — a simple indicator of satisfaction and usefulness.[19] When conversations are complete, certified coaches score random samples of conversations against corresponding rubrics, and those results are aggregated to ensure that Coach is meeting the bar for quality.[20]

Over time, the CareerVillage team has also built an AI model that can automatically evaluate Coach's responses against the given rubrics. This additional layer of evaluation means that every time a student uses Coach, dozens of evaluations are running on the conversation to assess the quality of Coach's responses.[21] Later, those evaluations will be matched against human scorers' results to gauge validity. This hybrid approach — human experts establishing benchmarks and AI scaling the evaluation process — allows CareerVillage to both maintain high standards for quality and feel confident that Coach is meeting them.

Predefined benchmarks for quality help CareerVillage optimize AI usage for flexibility and performance.

Coach, like many AI-powered ed tech tools, uses application programming interfaces (APIs) that connect to one or more frontier large language models (LLMs), including OpenAI's ChatGPT, Anthropic's Claude, and Google's Gemini, among others. While some tools may only use one model from one provider (e.g., ChatGPT-4o from OpenAI), Coach uses many models. As Chung said, "If you've heard of it, we've probably tried it."[22]

This approach introduces extra engineering and cost complexities since every model and provider has different usage rates; however, there are advantages. One is that using multiple models creates redundancy that reduces vulnerability to outages or downtime from any single provider. Chung also wanted his team to be able to select whichever model performed best for the activity at hand.[23] Rather than being limited to one provider or model, Coach developers can optimize for the highest-quality responses, as determined by the existing rubrics.

CareerVillage's benchmarks and evaluations also help the team balance performance with practical constraints. While the most advanced models might offer impressive answers, they are often slower than nonreasoning models and cost more for every token sent and received. The costs add up as chats get longer and Coach learns more about a student; any additional context must be continually revisited by the model for each conversation. Ultimately the developers found that the newest, most advanced models were often unnecessary. Given that activities were predetermined and users are typically younger or earlier in their career journeys, older models could provide just as high-quality answers while being cheaper and faster.[24] Having the predetermined, human-developed benchmarks for performance helped the CareerVillage team right-size their operations rather than simply relying on AI providers' latest developments.

Amplifying Learning

Career development is sometimes framed as more of a checklist of milestones than a learning journey — teachers are not necessarily testing for mastery on something like drafting a résumé. Consequently, using AI for career exploration may not bring up the same fears of cheating or cognitive offloading as other education use cases. CareerVillage therefore could have built a tool that emphasized fun exploration activities or time-saving techniques, but the Coach developers instead designed their AI tool with authentic, effortful learning in mind. To help students cultivate healthy cognitive mindsets, there are a few ways Coach introduces productive struggle to help students build psychological foundations for success:

1. **Conversations are tailored, not targeted, for healthy engagement based on students' interests.** Keeping any topic interesting is critical to capturing students' sustained attention, and it becomes even more critical in career development, where students may not see payoffs for years. At the same time, technology that targets only engagement can create harmful dependencies or hinder learning.[25] CareerVillage addressed this issue by tailoring Coach conversations to students' self-declared interests but keeping the application's interface simple. Chung noted that younger students might even find the interface boring,[26] but in the end, the more effortful engagement is likely to yield deeper learning.[27]

2. **Activities cultivating self-concept also build intrinsic motivation.** Self-concept refers to the beliefs students might hold about themselves. Positive beliefs can build students' intrinsic motivation by giving them a sense of agency when it comes to shaping their futures, but that intrinsic motivation is strongest when it comes from solving hard problems.[28] Coach simulates this using predefined activities: For example, rather than just sharing information about a career, Coach guides the learner through an activity so they explore and assess its fit from their own perspective.

3. **Practicing with feedback boosts metacognitive practices and self-efficacy.** Although linked to self-concept, CareerVillage distinguishes self-efficacy as a student's belief specifically about their capability of achieving a task or goal. As students engage with Coach and receive feedback on their practice, they calibrate their self-assessment and develop greater awareness of their own capabilities and areas for growth. These metacognitive practices, combined with practice in concrete skills, boost students' confidence in their ability to solve problems and achieve their goals.[29]

CareerVillage's team also made intentional choices about where not to use AI. In Coach's impact model, the team acknowledged that certain aspects of career-ready mindsets — such as belonging, curiosity, or mentoring relationships — require human connection that technology cannot replicate. This was validated during the pilot with Generation.org; despite users' overall optimism, surveys still revealed that "there is truly no replacement for face-to-face mentorship."[30] This finding led to a new feature where Coach, when applicable, will cite or direct students to answers from CareerVillage's community forums, where they can connect with real working professionals alongside their AI-assisted skill building.

Conclusion

CareerVillage continues to develop Coach's capabilities in ways that deepen student engagement and support. On the horizon is a more proactive version of Coach that reaches out to learners about what they are enjoying and learning, rather than waiting for students to initiate every interaction. The team is also developing specialized pathways that guide students through more comprehensive career exploration and preparation sequences.[31]

As AI-powered ed tech tools proliferate, Coach by CareerVillage demonstrates how thoughtful design grounded in research can move beyond simple question-answering toward meaningful skill and mindset development. By meeting learners outside the classroom and providing personalized practice opportunities at scale, Coach addresses real gaps in career development infrastructure without trying to replace human mentorship and connection. The organization's commitment to evidence-based impact measurement, rigorous quality evaluation, and clear-eyed assessment of AI's limitations offers an exemplar for others seeking to use technology to expand students' access to career support and information. ✦

Endnotes

1 "Our Mission," CareerVillage.org, https://www.careervillage.org/about/mission/.

2 "Complete List of Coach Activities," Coach by CareerVillage.org, https://www.aicareercoach.org/activities.

3 "School Counselor Roles and Ratios," American School Counselor Association, https://www.schoolcounselor.org/About-School-Counseling/School-Counselor-Roles-Ratios.

4 "School Counselors Matter," Education Trust, February 2019, https://www.schoolcounselor.org/getmedia/b079d17d-6265-4166-a120-3b1f56077649/School-Counselors-Matter.pdf.

5 Sarah Shaw and Abigail Lupi, *Evaluating Career Development Outcomes for Coach, a GenAI Project by CareerVillage.org* (CareerVillage.org, August 2024), 6, https://www.aicareercoach.org/measuring-career-readiness.

6 Shaw and Lupi, *Evaluating Career Development Outcomes for Coach, a GenAI Project by CareerVillage.org*, 6.

7 Shaw and Lupi, *Evaluating Career Development Outcomes for Coach, a GenAI Project by CareerVillage.org*, 6.

8 Shaw and Lupi, *Evaluating Career Development Outcomes for Coach, a GenAI Project by CareerVillage.org*, 6.

9 Shaw and Lupi, *Evaluating Career Development Outcomes for Coach, a GenAI Project by CareerVillage.org*, 7.

10 "Impact," Coach by CareerVillage.org, https://www.aicareercoach.org/impact.

11 "Home," EDSAFE AI Industry Council, https://www.edsafeai.org/industry-council.

12 Jared Chung, interview, June 23, 2025.

13 "Quality and Accuracy," Coach by CareerVillage.org, https://www.aicareercoach.org/quality-assurance.

14 "Safety and Trust," Coach by CareerVillage.org, https://www.aicareercoach.org/safety.

15 Jared Chung, interview, June 23, 2025.

16 Ibid.

17 "Quality and Accuracy," Coach by CareerVillage.org.

18 Jared Chung, interview, June 23, 2025.

19 Ibid.

20 Ibid.

21 Ibid.

22 Ibid.

23 Ibid.

24 Ibid.

25 Kevin Fuchs, "Challenges with Gamification in Higher Education: A Narrative Review with Implications for Educators and Policymakers," *International Journal of Changes in Education* 1, no. 1 (2024), https://doi.org/10.47852/bonviewIJCE32021604.

26 Jared Chung, interview, June 23, 2025.

27 Amy Chen Kulesa, Marisa Mission, Michelle Croft, and Mary K. Wells, *Productive Struggle: How Artificial Intelligence Is Changing Learning, Effort, and Youth Development in Education* (Bellwether, June 2025), https://bellwether.org/publications/productive-struggle/.

28 Kulesa, Mission, Croft, and Wells, *Productive Struggle: How Artificial Intelligence Is Changing Learning, Effort, and Youth Development in Education*; Kenneth W. K. Lo, Grace Ngai, Stephen C. F. Chan, and Kam-por Kwan, "How Students' Motivation and Learning Experience Affect Their Service Learning Outcomes: A Structural Equation Modeling Analysis," *Frontiers in Psychology* 13 (April 2022), https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.825902/full.

29 Kulesa, Mission, Croft, and Wells, *Productive Struggle: How Artificial Intelligence Is Changing Learning, Effort, and Youth Development in Education*.

30 Shaw and Lupi, *Evaluating Career Development Outcomes for Coach, a GenAI Project by CareerVillage.org*, 7.

31 See for example: "Building AI-Powered Career Verticals with University of Florida College of Nursing," Coach by CareerVillage.org, https://www.aicareercoach.org/case-studies/uf-nursing.

About the Authors

MARISA MISSION

Marisa Mission is a senior analyst at Bellwether. She can be reached at **marisa.mission@bellwether.org**.

MICHELLE CROFT

Michelle Croft is an associate partner at Bellwether. She can be reached at **michelle.croft@bellwether.org**.

AMY CHEN KULESA

Amy Chen Kulesa is a senior associate partner at Bellwether and leads the organization's work on AI. She can be reached at **amy.chenkulesa@bellwether.org**.

About Bellwether

Bellwether is a national nonprofit that works to transform education to ensure young people — especially those furthest from opportunity — achieve outcomes that lead to fulfilling lives and flourishing communities. Founded in 2010, we help mission-driven partners accelerate their impact, inform and influence policy and program design, and bring leaders together to drive change on education's most pressing challenges. For more, visit **bellwether.org**.

ACKNOWLEDGMENTS

We would like to thank the experts who gave their time and shared their knowledge with us to inform our work, including Jared Chung and CareerVillage staff. Thank you also to the Bezos Family Foundation, Chan Zuckerberg Initiative, Charter School Growth Fund, and Overdeck Family Foundation for their financial support of this project.

We would also like to thank our Bellwether colleagues Mark Baxter and Mary K. Wells for their input and former colleague Janine Sandy for her support. Thank you to Amy Ribock, Kate Stein, Andy Jacob, McKenzie Maxson, Temim Fruchter, Julie Nguyen, and Amber Walker for shepherding and disseminating this work, and to Super Copy Editors.

The contributions of these individuals and entities significantly enhanced our work; however, any errors in fact or analysis remain the responsibility of the authors.