Students Choose Human Counselors Over Algorithms in College Applications, but Not Always*

Hemanshu Das Sofoklis Goulas Faidra Monachou

April 18, 2025

This is a preliminary draft; please do not distribute.

Abstract

This study examines student preferences for human versus algorithmic recommendations in college applications. Conducted across 14 public high schools in Greece, the experiment reveals that students exhibit aversion to algorithmic recommendations when the recommendation basis is more objective but not when it is most subjective. We find that student perceptions of the recommender's intent strongly drive this aversion, consistently across scenarios and statistical approaches; perceptions of alignment with personal goals, ability, and comprehension also play significant roles. The results further reveal substantial heterogeneity in recommendation adoption rates across several dimensions, including gender, academic performance, adherence to the norm of "prestige chasing," and school urbanicity. Free-text student responses suggest that students seek guidance and information about alternative study options from human counselors but turn to algorithms for recommendations based on grades and admissions chances. Using an optimization approach, we demonstrate how a planner can navigate the heterogeneity in recommendation adoption rates and optimally prioritize the assignment of human versus algorithmic recommenders, under varying social preferences and limited capacity of human counselors. We find that a targeting policy relying on few readily available student and school features can approximate the first-best, personalized targeting policy effectively. These insights underscore the importance of understanding student preferences in designing effective and equitable recommendation systems, and highlight the potential of hybrid approaches that integrate human guidance with algorithmic tools.

Keywords: recommendations; college applications; algorithm aversion; AI in education; experiment.

* *Das*: Yale University, School of Management (hemanshu.das@yale.edu); *Goulas*: Brookings Institution, Economic Studies, and IZA (sgoulas@brookings.edu); *Monachou*: Yale University, School of Management (faidra.monachou@yale.edu). The study was IRB-approved through Yale University with approval identification number 2000038360 (in coordination with the Regional Directorate of Primary and Secondary Education and following the guidelines of the national Institute of Education Policy) and pre-registered in the AEA RCT Registry with unique identification number AEARCTR-0014368. The authors would like to thank Manos Alevras for his assistance with text data digitization. The authors are grateful to Adriana Germano for helpful discussions and suggestions.

1 Introduction

The integration of revolutionary AI (Artificial Intelligence) technology in education—ranging from personalized learning to curriculum development to college admissions (Claybourn, 2023)—has sparked a wave of ambitious initiatives, driven by the belief in its transformative potential (Spector, 2024). Policymakers around the world are increasingly recognizing this potential and are actively working to support the adoption and integration of educational technology solutions; see, e.g., the report by the Connecticut Commission for Educational Technology (2024). As Saavedraez and Molina (2024) at the World Bank note, AI is *"probably the most significant transformation in education since the printing press."*

Despite significant investments in AI for education, adoption is often slowed by low trust and familiarity with AI, along with concerns regarding its impact on effectiveness and equity (Varsik and Vosberg, 2024). Countries like the US and Greece—where our study is based—are actively exploring how best to integrate AI into schools. In the US, most states remain cautious in offering AI guidance at the school and district levels (Dusseault and Lee, 2023), and AI's role in high-stakes decisions like grading and admissions remains minimal due to fairness concerns (Force, 2024). Greece has launched a national initiative to bring AI tools into classrooms (AlfaVita, 2024) including curriculum changes and pilot programs (Kathimerini, 2023, 2024). However, student familiarity with AI is high: over 70% of middle and high school students report using large language models such as ChatGPT for schoolwork (Sidoti et al., 2025; Zhu et al., 2024), and algorithm-driven platforms are deeply embedded in adolescent culture (Kang and Lou, 2022; Tsitsika et al., 2014). The ultimate challenge lies not in the ability to create advanced tools—AI's boundless capabilities are well-established—but in integrating them into practice in ways that resonate with students and educators, ensuring their effectiveness and successful adoption.

Recommendation algorithms, for example, hold immense potential to improve educational decision-making by guiding students through complex choices, such as college applications (Gedrimiene et al., 2023). Research has documented that lack of awareness about available options and biases in beliefs about admission chances can lead students to suboptimal decisions (Arteaga et al., 2022; Bobba and Frisancho, 2022; Larroucau et al., 2024). Thus, AI-driven recommendations can aid decision-making by leveraging massive amounts of data and computational power to process complex information. Furthermore, AI has the potential to help students at scale, democratizing access to the costly expertise of human counselors. This is particularly valuable in typically resource-constrained environments such as public education systems.

However, the effectiveness of such AI-based recommendations may be limited if students resist their use. Indeed, *algorithm aversion*, the phenomenon where individuals are reluctant to adopt advice from algorithms, even when these algorithms consistently outperform human judgment (Dietvorst et al., 2015), has been observed across several domains, such as medical decision-making (Lin et al., 2021; Longoni et al., 2019) and consumer preferences (Vodrahalli et al., 2022). Nevertheless, this phenomenon has yet to be explored in the context of educational decision-making, particularly in high-stakes college admissions.

This paper studies how students respond to AI-based algorithmic recommendations versus those from human counselors in the college application process. We conducted a lab-in-the-field, survey-based experiment across 14 public high schools in Greece. Similarly to 40% of countries worldwide (Neilson, 2024), Greece has a centralized system for higher education, utilizing a standardized national exam and a centralized college admissions process. The experiment compared the adoption rates of identical recommendations provided by algorithms and human counselors across scenarios that varied in the objectivity of the recommendation basis. We define an objective recommendation basis as one grounded in quantifiable and measurable facts with minimal ambiguity, whereas a subjective recommendation basis may reflect personal opinion (Castelo et al., 2019). We examined three hypothetical, third-person scenarios. In particular, the scenarios test the prestige of a program against students' personal interests (*Heart* scenario), geographic location (*Geography* scenario), and admissions chances (*Pragmatism* scenario). These scenarios reflect varying degrees of objectivity of their recommendation basis, with *Heart* having the most subjective and *Pragmatism* having the most objective.

Our results reveal significant student aversion to algorithmic recommendations compared to those from human counselors when the recommendation basis is more objective. We observe the largest and statistically significant adoption rate gap of 4.7 percentage points in the most objective scenario (*Pragmatism*). However, when recommendations are based on the most subjective criterion (*Heart*), algorithmic recommenders perform similarly to human counselors. Our analysis confirms this general pattern across multiple dimensions, including gender, prior-year academic performance, adherence to the norm of "prestige chasing", school locale, grade level, intention to participate in the national exams, and track choice.

Recommendation adoption rates exhibit substantial heterogeneity by various student and school characteristics. Regarding gender, an interesting reversal occurs as the objectivity of the recommendation basis increases. While female students adopt recommendations at a lower rate than male students in more subjective scenarios (*Heart* and *Geography*), the pattern flips in the most objective scenario (*Pragmatism*). Here, female students show larger adoption rates than male students, with an 11.0 percentage-point difference for human counselors and a 12.2 percentage-point difference for algorithmic recommenders between the descriptive adoption rates of female and male students. Consistent with our main result, both female and male students show the greatest algorithm aversion in the most objective scenario (*Pragmatism*), but male students exhibit greater aversion than female students, although the difference is not statistically significant. Overall, although not statistically precise, algorithm aversion persists across all scenarios for male students. The pattern is different for female students. Specifically, in the most subjective scenario (*Heart*), female students marginally prefer algorithmic recommenders over human counselors, whereas male students continue to exhibit weakly significant algorithm aversion.

Prior academic performance, school urbanicity, and compliance to the social norm of "prestige-chasing" are further associated with heterogeneity in adoption rates. With respect to prior academic performance, as reflected in the students' prior-year GPA, we find a mismatch in the preferences of high- and low-achieving students. High-achieving students show no preference in the most objective scenario (*Pragmatism*). In contrast, low-achieving students significantly prefer human counselors in pragmatic decision-making, with a significant 8.0 percentage-point gap. High-achieving students, on the other hand, are more likely to prefer human recommendations in *Geography*, with a significant 6.9 percentage-point adoption rate gap. With respect to school locale, we find that rural participants are more sensitive to the source of recommendations in the objective scenario (*Pragmatism*), significantly favoring human over algorithmic recommendations with a 9.3 percentage-point gap. In contrast, urban participants show no significant preference between recommendation types in any scenario. Moreover, we measure compliance to the norm of "prestige-chasing" by the students' intention to pursue the most prestigious academic programs they can gain admission to, regardless of their personal preferences. On average, students self-report low norm compliance, but the variance in responses is high. Our results show that higher norm compliance scores are associated with higher aversion to algorithmic recommenders in the two more objective scenarios (*Geography* and *Pragmatism*).

To better understand how to overcome these adoption barriers, we further investigate the mechanisms behind students' aversion to algorithmic recommendations. Specifically, we examine four mediators that may explain why students resist algorithmic recommendations. These mediators include perceived *intent* (whether the recommender is perceived as acting in the persona's best interest), *ability* (confidence in the recommender's competence in providing helpful recommendations), *comprehension* (clarity of the recommendation basis's reasoning), and *alignment* (how well the recommender's suggestion aligns with the persona's study preferences).

Our results reveal that all four mediators influence recommendation adoption across different scenarios in varying degrees. Most notably, *intent* significantly mediates students' aversion to algorithmic recommendations, consistently across all scenarios and statistical tests performed. Its effect remains significant when we control for the other mediators across all scenarios. In the most objective scenario (*Pragmatism*) that also has the highest observed algorithmic aversion, *alignment* is found to be a statistically significant mediator, while its explanatory contribution is the largest when we control for other mediators. Combining multiple mediators increases explanatory power, suggesting they capture overlapping dimensions of decision-making.

To get insights on the reasons behind students' algorithm aversion that structured responses on adoption mechanisms alone may not fully reveaal, we further collected and analyzed free-text responses. Using a transformer-based natural language inference model, we classified students' text responses. We find that students perceive human counselors as more helpful than algorithmic recommenders (a score of 0.78 for human counselors vs. 0.67 for algorithmic recommenders on helpfulness). For example, a participant in our study noted in an open-ended question: "I would not trust an algorithm for such an important decision [college applications]." Our findings indicate that students seek guidance and exploration of program options from human counselors but interestingly turn to algorithmic recommenders for decision support based on grades and admissions probabilities, despite their observed aversion to algorithms.

In a policy application, we use our insights to optimally allocate human counselors versus algorithms to students. While our experiment showed that human counselors are generally preferred over algorithms, they are costly and scarce, particularly in resource-constrained public schools. Moreover, the heterogeneity in adoption rates of algorithmic versus human recommendations among different student subgroups presents an opportunity for optimized targeting. In our policy application, we adopt a social planner's perspective to differentiate which students receive human counselors and which receive algorithmic recommendations, with the goal of improving the overall recommendation adoption. We use an optimization approach to find the first-best (personalized) targeting policy as well as approximation solutions relying only on a few readily available features. We further take into account that planners may vary in their attitudes towards students' adoption rates; for example, a planner may prefer to allocate human counselors to students who are most resistant to algorithmic recommendations or to those who are already receptive to guidance from any source.

We find that prioritizing students with the highest baseline (algorithm) adoption leads to first targeting female students in the Life Sciences track at rural schools, whereas focusing on those with lower baseline adoption directs resources first to male students in the Exact Sciences track at rural schools. To reduce complexity and costs, planners may also consider implementing simpler targeting policies based on fewer student and school characteristics. We find that a targeting policy using four student and school characteristics—urbanicity, gender, track, and prior-year GPA—achieves 87.6% of the adoption gains made by the first-best (personalized) targeting policy. Our policy exercise highlights that targeted policies can yield significant benefits at manageable complexity, if they rely on the appropriate student characteristics.

Overall, our paper makes several novel contributions. To the best of our knowledge, this is the first study to explore algorithm aversion in the educational context, particularly in high-stakes decision-making (Castelo et al., 2019; Dargnies et al., 2024; Sunstein and Gaffe, 2024), among a new population—adolescents. Second, our results shed light on various novel mechanisms behind algorithm aversion. We find that aversion is strongly driven by perceptions of the *intent* of the recommender, as well as by *alignment* in the most objective scenario, followed by *ability* and *comprehension* to a lesser, yet significant degree. Additionally, we identify substantial heterogeneity in aversion based on the subjectivity of the recommendation basis, student gender, academic performance, norm compliance, and school urbanicity. Notably, our study is also the first to investigate gender differences in algorithm aversion, emphasizing the critical role these differences play in recommendation adoption. In a policy application motivated by the practical challenge planners face in allocating costly resources like career counselors, we apply our experiment insights to a targeting problem. We take an optimization approach to determine the optimal allocation between human counselors and algorithmic recommenders, considering varying planner priorities. Our policy insights demonstrate how planners can allocate resources more effectively and cost-efficiently, following a hybrid human-algorithm targeting approach. More broadly, our study contributes to the effective design of algorithmic systems and highlights the need for incorporating human-like features in AI systems.

1.1 Literature Review and Hypothesis Development

Algorithm Aversion and Algorithm Appreciation. Resistance to algorithmic systems has been widely documented across contexts, including medical decision-making (Cadario et al., 2021; Lin et al., 2021; Longoni et al., 2019), market demand forecasting (Dietvorst et al., 2015), fraud detection (Boatsman et al., 1997), crime mitigation (Kleinberg et al., 2017), recruitment (Highhouse, 2008), administrative decision-making in higher education (Xu et al., 2023), and consumer preferences (Vodrahalli et al., 2022; Yeomans et al., 2019). Newman et al. (2020) and Dargnies et al. (2024) further investigate the cognitive biases that exacerbate resistance, such as overconfidence in human judgment and the illusion of control. Recent studies, such as Xu et al. (2023), also explore interventions to mitigate algorithm aversion, including providing users with explanatory feedback and enabling human oversight of AI systems.

While much research has focused on aversion to algorithms, individuals may demonstrate a preference for algorithmic recommendations in contexts where algorithms are perceived as objective and consistent (Vodrahalli et al., 2022). For example, Logg et al. (2019) show that algorithmic appreciation is more likely in tasks requiring expertise and objectivity.

Trust in AI and Drivers of Recommendation Adoption: Task characteristics significantly influence trust in AI. A body of work (Castelo et al., 2019; Glikson and Woolley, 2020; Logg et al., 2019; Sunstein and Gaffe, 2024; Vodrahalli et al., 2022; Yeomans et al., 2019) finds that people are more likely to trust

AI in objective, quantitative, or repetitive tasks, but favor human input in subjective tasks requiring intuition or empathy. Conversely, Lee (2018) finds that people perceive algorithms as lacking subjectivity, making them less likely to trust algorithmic recommendations if they perceive that the scenario requires a subjective evaluation (Bogert et al., 2021; Longoni and Cian, 2020).

The adoption of recommendations is influenced by various drivers, such as expectations and expertise, decision autonomy, incentivization, cognitive compatibility, and divergent rationalities (Burton et al., 2020). *Intent* plays a critical role, as users are more likely to adopt recommendations if they perceive the system as prioritizing their benefit over profit or undue influence (Komiak and Benbasat, 2006). Confidence in the algorithm's competence (*ability*) is a stronger predictor of recommendation adoption than confidence in its good faith (Choung et al., 2023; Prahl and Van Swol, 2017; Zerilli et al., 2022). *Comprehension* enhances adoption when users understand the rationale behind recommendations. Transparency, explainability, and control over the recommendation process build trust and enable users to integrate automated suggestions while maintaining autonomy (Bartmann, 2023; Kizilcec, 2016). Finally, *alignment* with users' initial beliefs fosters adoption, as users tend to accept recommendations closer to their existing views, providing validation and reinforcing confidence in their judgments (Bonaccio and Dalal, 2006). However, individuals often prefer to exercise agency and stick to the default choice. When the default choice is different than the recommendation, and the recommender does not provide sufficient explanation, there might be a lack of perceived *alignment* (Shoval et al., 2022).

Operations for Educational Policy and EdTech. Applications of operations research methods to education policy have gained increasing attention in recent years (see, e.g., Allman et al. 2022; Arteaga et al. 2022; Bastani et al. 2024; Faenza et al. 2020; Garg et al. 2020; Goulas and Monachou 2024; Goyal et al. 2024; Keppler et al. 2022; Smilowitz and Keppler 2020). A subset of these works has focused on EdTech, with studies examining the impact of generative AI on students' learning (Bastani et al., 2024) and teachers' backwards planning (Keppler et al., 2024), optimizing parental engagement (Goyal et al., 2024), crowdfunding for teachers (Keppler et al., 2022), and technology-based nudges in school choice (Arteaga et al., 2022). Our work contributes to this growing field by combining experimental methods with an optimization approach, and to the best of our knowledge, it is the first to explore the use of algorithmic recommendations in college applications.

Hypothesis Development

The existing literature has largely overlooked the specific contexts of college admissions and educational technology (EdTech), as well as the unique decision-making characteristics of adolescents. On the one hand, adolescents may be more inclined to use technology (Gibbons and Poelker, 2020), but they are also still developing the metacognitive ability to accurately assess the quality of their own knowledge and make independent decisions (Moses-Payne et al., 2021). Moreover, factors such as gender and other demographic variables remain underexplored in this area. Motivated by this gap and building upon the previous literature, we now propose hypotheses regarding the adoption of algorithmic recommenders versus human counselors in the context of college applications:

Hypothesis 1a *Students prefer human counselors over algorithmic recommenders for college application recommendations.*

Hypothesis 1b *Students prefer algorithmic recommenders over human counselors for college application recommendations.*

Given the role of task subjectivity in adoption of recommendations (Castelo et al., 2019), we intend to test our hypothesis across three scenarios of college application recommendations, each varying in the objectivity of the recommendation basis. The first two scenarios will require more subjective judgment, focusing on geographic preferences and personal interests. The third scenario will present a rather objective recommendation criterion, based on admission chances and grades. Further details on the scenarios are provided in Section 2.3.

2 Experimental Design

This section outlines the education system in Greece and describes the experiment's design, treatment conditions, and implementation procedure.

2.1 The Education System in Greece

The Greek education system provides a highly standardized setting for studying college application behavior. It is characterized by universal access to college education and identical infrastructure and funding across schools (Organisation for Economic Co-operation and Development, 2018). Over 90% of high school students in Greece attend public schools, with grade 10 marking the beginning of high school. For grades 11 and 12, students select one of four specialization tracks: *Economics and Information Technology* (IT), *Exact Sciences*, *Humanities*, or *Life Sciences*. These tracks are universally available in all schools, and students must finalize their selection before the start of grade 11. Each track requires students to complete a distinct set of courses to graduate.[1] The selected track also determines the set of degree programs to which students can apply for admission. For example, students in the Humanities track are eligible for humanities-related degree programs but cannot apply to programs in engineering or health sciences.

College education in Greece is tuition-free, and the admissions process is highly centralized. Nearly all tertiary students in Greece attend public universities (Eurostat, 2024). At the end of grade 12, students take externally graded and proctored national exams, which serve as the primary determinant for college admissions. After the exams, students submit a ranked list of preferred postsecondary programs to the Ministry of Education. A centralized, computerized system then ranks students nationwide by their exam scores and sequentially assigns each to their highest-ranked program with available seats, based on the fixed admission capacities determined annually by the Ministry. Submitting a list of preferences incurs no financial cost, and students receive only one admission offer, corresponding to their highest-ranked program for which they qualify. At the end of the process, each department announces its "admission threshold score," which reflects the score of the lowest-ranked student admitted that year. These thresholds are shaped by the demand for specific programs, as indicated by students' top-choice applications and the available supply of seats. Programs with higher admission threshold scores are typically perceived as more prestigious.

[1] While students have the option to remain in their chosen track or switch tracks between the end of grade 11 and the beginning of grade 12, fewer than 1% make such changes (Goulas et al., 2024).

2.2 Experimental Setting

The experiment was conducted in 14 public high schools in Greece.[2] These schools were selected in collaboration with a Regional Directorate of Primary and Secondary Education ("RDE"), which oversees all public schools in the region, serving more than 100,000 students. The selection process aimed to ensure a representative sample of schools based on location, size, and socioeconomic factors, thereby providing a diverse and balanced participant pool. All schools included in the study were regular day schools, excluding special education institutions.

The target population consisted of students in grades 10, 11, and 12. These students are already familiar with career orientation and university track choices, as Greek high schools play a key role in guiding students through the college admissions process. Table S1 compares the characteristics of the sampled schools to the overall population of high schools in Greece. The sampled schools represent a diverse range of student and school characteristics and align closely with the population across all variables. Key characteristics such as gender distribution, student age, track choice proportions, STEM application rates, and university admission scores are broadly similar between the sample and the population. The proportion of students admitted to higher education institutions is slightly lower in the sampled schools (53% vs. 57%), but this difference is not statistically significant. Moreover, the sampled schools are statistically comparable to the population in terms of the share of rural schools and the average annual income of their locations. Overall, the sampled schools provide a reasonably representative subset of Greek high schools, with minor differences likely reflecting regional variation in socioeconomic contexts.

2.3 Treatments

The experiment involves two treatment conditions: (1) recommendations for college applications provided by a **human counselor**, and (2) recommendations for college applications provided by an **algorithmic recommender**. These treatments investigate differences in students' perceptions and adoption rate based on the source of the recommendations.

Participants were randomly assigned to one of the two conditions and completed a survey featuring three hypothetical third-person scenarios involving fictional students making college application decisions. These scenarios were chosen to address the absence of real-world algorithmic recommenders for college admissions while mitigating the limitations of hypothetical, self-referential scenarios. Third-person scenarios offer less abstraction, greater control over situational details, and reduce the likelihood of eliciting aspirational rather than actual behavioral responses (Logg and Schlund, 2024).

The scenarios were as follows:

1. **Prestige vs. Personal Interests** (*Heart*): The recommendation encouraged the student to pursue a prestigious program over prioritizing their own personal academic and career interest. ~~The recommendation discouraged the student from prioritizing their personal academic and career interests over pursuing a prestigious program.~~

2. **Prestige vs. Geographic Location** (*Geography*): The recommendation encouraged the student to prioritize a program closer to home over a more prestigious college program in the capital city.

[2] The study was IRB-approved through an affiliated university and pre-registered in the AEA RCT Registry.

3. **Prestige vs. Admission Chances** (*Pragmatism*): The recommendation encouraged the student to apply to an attainable program with lower prestige to improve their admission chances.

The experiment aimed to explore how students respond to recommendations from human counselors versus algorithmic systems. Except for the recommendation source, all scenario details were identical across treatment conditions. Each scenario featured a female student deciding how to rank specific degree programs on her application form. Between varying the persona's gender or not, we chose the latter to ensure sufficient statistical power and comply with external logistical constraints. Each scenario included a brief profile describing the student's national exam performance, personal circumstances, or preferences to provide context and plausibility, along with her preliminary ranking of degree programs. The competitiveness of each program was indicated by its prior-year admission cutoff (i.e., the national exam score of the last admitted student). A recommendation was then presented, along with the criterion used to generate it. For example, in the first scenario, the recommender prioritized proximity to the student's home location. To ensure students had sufficient background knowledge to understand the scenarios, the bottom half of the first page of the survey—before the scenarios—provided information about the national exams (Panelladikes) and the college application and admission process.

The recommendation criteria across the three scenarios vary in objectivity. We consider an objective recommendation basis as one grounded in quantifiable and measurable facts with minimal ambiguity, whereas a subjective recommendation basis may reflect personal opinion (Castelo et al., 2019).The *Heart* scenario has the least objective recommendation basis, as prioritizing a popular field over a student's explicitly stated passion may seem subjective, arbitrary, or even unreasonable. Encouraging students to pursue a popular field with little regard for personal interests is less likely to resonate with many students. In contrast, the *Pragmatism* scenario seems the most objective, as basing recommendations on admission chances is widely perceived as rational, commonsensical, and unambiguous; however, it may also appear impersonal and lacking empathy. The *Geography* scenario falls somewhere in between: recommending studying close to home may be seen as practical due to cost savings but could also appear arbitrary or at odds with a student's desire to experience new places. Prior research suggests that algorithmic aversion is less likely in objective tasks (Vodrahalli et al., 2022; Yeomans et al., 2019), while users often prefer human advice for tasks requiring subjective judgment (Longoni et al., 2019); Castelo et al. (2019) reports mixed results. Nevertheless, it remains unclear how the objectivity of the recommendation basis influences recommendation adoption. The variation across the three scenarios in this study allows us to explore how students respond to different levels of objectivity in the recommendation criteria.

Moreover, we varied how well the subjective recommendations aligned with students' personal study preferences in the first two scenarios. Prior research suggests that users are more likely to adopt algorithmic recommendations when they feel personalized (Dietvorst et al., 2018). In the *Geography* scenario, the persona expresses no explicit preference for studying near home, whereas in the *Heart* scenario, the persona is passionate about the field that the recommender discourages. This contrast allowed us to examine whether acknowledging and validating students' preferences influences their acceptance of algorithmic advice.

The survey design included measures for primary outcomes and secondary outcomes (mediators). The primary outcome was *recommendation adoption*, measured as a binary variable indicating whether the participant accepted the recommendation or not. The mediators included trust in *intent* (whether the recommender is perceived as acting in the persona's best interest), perceived *ability* (confidence in the recommender's competence in providing helpful recommendations), rationale *comprehension* (clarity of the recommendation basis's reasoning), and preference *alignment* (how well the recommender's suggestion aligns with the persona's study preferences). Mediators were measured using agreement scores on a Likert-type scale ranging from -5 (Completely Disagree) to +5 (Completely Agree). To limit bias, questions related to mediators appeared before the primary outcome question regarding recommendation adoption. By comparing responses across the treatment groups, the study identifies systematic differences in students' perceptions of human versus algorithmic recommendation systems in a centralized college admissions context. The full questionnaires can be found in Appendix Section S1.

2.4 Experimental Procedure

Prior to the school visits, parental consent forms were distributed and collected with the assistance of school administrators, while student assent forms were distributed and collected by the research team on the day of the survey. On the day of the survey, participating students were randomly assigned to one of two treatment conditions—evaluating recommendations from either a human counselor or an algorithmic recommender—through the distribution of shuffled paper-based surveys, ensuring an even allocation across treatments. During the experiment, participating students remained in their designated classrooms. The surveys were completed anonymously in approximately 15 minutes, collected by the research team, and digitized for analysis. The survey design, administration process, and classroom setting ensured that data collection was consistent across schools and minimally disruptive to regular school activities.

Table 1 provides summary statistics of participant characteristics across the two treatment conditions: human counselors and algorithmic recommenders. Panel A indicates that prior academic performance is comparable between groups, with no significant differences observed in GPA or subject scores. Panel B shows that the majority of participants are female,[3] and approximately one-third of participants are from each grade level. Most participants (over 83%) intend to take the national exams. Track choice distributions are balanced across conditions. Panel C explores norm compliance, measured by agreement with the statement, *"I will apply to the most competitive programs I have a chance of being admitted to regardless of my other interests."* The average scores on this scale (ranging from -5 to 5) reflect low agreement with the statement, indicating low average norm compliance, with no significant differences between the treatment groups; however, the high standard deviation in the responses (2.8-2.9, almost twice the mean) indicates great variability in students' views as a whole. Overall, the two treatment groups are well-balanced, supporting the validity of comparisons between the human counselor and algorithmic recommender treatments.

2.5 Identification

We estimate the impact of human versus algorithmic recommendations on the likelihood of recommendation adoption using the following specification:

$$\mathbb{P}(Adoption_{is} = 1) = F(\alpha + \beta Algorithmic\ Recommender_i + \gamma \mathbf{X}_i + \eta_s + \epsilon_{is}) \tag{1}$$

[3] Female students are generally more likely to respond to survey requests, a tendency linked to their greater pro-social behavior (DellaVigna et al., 2013).

Table 1: Participant Characteristics

	Human Counselor			Algorithmic Recommender				
Panel A: Prior Performance								
	Mean	SD	N	Mean	SD	N	Diff.	p-value
GPA	17.2	1.9	1,052	17.1	2.1	999	-0.090	0.313
Mathematics	13.9	5.4	1,016	14.2	5.1	946	0.377	0.113
Greek Language	15.9	2.9	1,018	15.9	2.8	955	0.007	0.960
Panel B: Participant Characteristics								
	%		N	%		N	Diff.	p-value
Gender								
Female	62.3		1,064	59.0		1,009	-3.343	0.120
Male	35.3		1,064	38.7		1,009	3.314	0.118
Grade								
10	35.9		1,053	35.8		1,001	-0.133	0.950
11	33.1		1,053	36.2		1,001	3.020	0.151
12	31.0		1,053	28.1		1,001	-2.887	0.152
Track								
Humanities	28.9		1,042	27.3		993	-1.596	0.424
Life Sciences	20.5		1,042	20.7		993	0.208	0.908
Exact Sciences	24.5		1,042	23.2		993	-1.310	0.488
Economics and IT	26.1		1,042	28.8		993	2.698	0.173
National Exam Participation								
Yes	83.7		1,063	83.7		1,011	-0.046	0.978
No	4.6		1,063	3.7		1,011	-0.950	0.277
Undecided	11.7		1,063	12.7		1,011	0.996	0.488
Panel C: Norm Compliance								
"I will apply to the most competitive programs I have a chance of being admitted to regardless of my other interests."								
	Mean	SD	N	Mean	SD	N	Diff.	p-value
Score	-1.6	2.8	1,041	-1.5	2.9	981	0.095	0.462

Notes: Panel A presents the moments of students' prior performance and their differences across treatment conditions. *Grade Point Average* (GPA) refers to the overall GPA from the prior year. Prior performance in *Mathematics* and *Greek Language* refers to students' scores on the final exams from the prior year. Both scores and GPA range from 0 to 20. Panel B reports participant characteristics for each treatment condition, along with the differences between them. *Track* choice for grade 10 students refers to their intended choice, as only grade 11 and grade 12 students are formally assigned to tracks. *National Exam Participation* refers to each student's intention to participate in the national exams for university admission at the end of grade 12. Panel C reports moments of the norm compliance score. *Norm compliance* is measured on a scale from -5 to 5, based on students' agreement with the statement: *"I will apply to the most competitive programs I have a chance of being admitted to regardless of my other interests."* A total of 1,068 students were assigned to the *Human Counselor* treatment condition, while 1,014 students were assigned to the *Algorithmic Recommender* treatment condition. *N* represents the number of non-missing records within each characteristic group.

In this specification, $\mathbb{P}(Adoption_{is} = 1)$ denotes the probability that student i at school s adopts the recommendation provided. $Algorithmic\ Recommender_i$ is an indicator variable taking the value 1 for students assigned to the algorithmic recommender condition. Students assigned to the human counselor condition serve as the reference group. β is the parameter of interest and captures the difference in recommendation adoption rates between algorithmic recommenders and human counselors, holding all other variables constant; we refer to $\hat{\beta}$ as the estimated adoption rate gap. Vector $\mathbf{X}_i$ includes student-level covariates, such as gender, grade level, track choice, intention to take the national exams, and prior academic performance in GPA and key subjects. To minimize record loss, indicators for missing covariate information are also included. School fixed effects (η_s) are included in all specifications to address potential location-level unobserved heterogeneity. Specification (1) is estimated using a linear probability regression model. For robustness, we report similar estimates using a logistic regression model in the Appendix (see Table S2). Standard errors are clustered at the school level to account for potential within-school correlation in responses.

Heterogeneity analyses are conducted by replacing the treatment variable $Algorithmic\ Recommender_i$ in specification (1) with group-specific treatment indicators. These analyses examine differential adoption rate gaps based on gender, prior academic performance, norm compliance levels, and school locale (urbanicity). For academic performance, students are categorized into tertiles of their prior-year GPA. Similarly, we investigate heterogeneity by tertiles of norm compliance scores.

With random assignment, simple comparisons of adoption rates can identify the relative effects of recommendation sources. The main assumption for obtaining causal estimates of β in specification (1) is that no omitted variables are correlated with both treatment assignment and the adoption outcome. Table 1, which compares student characteristics across the human and algorithmic recommendation treatment groups, helps assess the validity of this assumption. The results indicate no statistically significant differences in characteristics such as gender, grade level, GPA, track choice, or intention to take the national exams. This balance ensures that random assignment successfully eliminates confounding factors, providing confidence that the treatment assignment is orthogonal to both observed and unobserved student characteristics.

3 Results

3.1 Main Estimates

Table 2 presents the estimated recommendation adoption rates (i.e., the proportion of students following recommendation over persona's original choice) across the three scenarios that differ in the objectivity of their recommendation criteria. As we move from the most subjective scenario (*Heart*) to the most objective one (*Pragmatism*), we find that the overall recommendation adoption rate increases, regardless of the recommendation source. Importantly, the adoption rate gap between human counselors and algorithmic recommenders also increases. Significant differences emerge in the *Geography* and *Pragmatism* scenarios, where adoption rates for algorithmic recommendations are lower by 3.0 and 4.7 percentage points, respectively ($\hat{\beta} = -0.030$, SE=0.016; $\hat{\beta} = -0.047$, SE=0.023). In contrast, the *Heart* scenario shows no significant difference between human and algorithmic recommendations ($\hat{\beta} = -0.008$, SE=0.023).

Our results indicate that algorithmic recommendation systems perform comparably to human counselors in subjective scenarios but face greater aversion when recommendations are based on more objective criteria. Interestingly, these findings partially contrast with prior work, such as Yeomans et al. (2019) and Vodrahalli et al. (2022), which show that higher task objectivity is associated with lower algorithm aversion, as well as Castelo et al. (2019) which reports mixed results on the role of objectivity. However, these studies have not considered the role of the recommendation basis' objectivity in algorithm aversion.

Table 2: Estimated Recommendation Adoption Rates

	Means		Without Controls		With Controls		
Scenario	Human	Algorithm	$\hat{\beta}$	SE	$\hat{\beta}$	SE	N
Heart	0.315	0.318	-0.003	0.024	-0.008	0.023	2,048
Geography	0.365	0.343	-0.030*	0.016	-0.030*	0.016	2,057
Pragmatism	0.471	0.419	-0.050**	0.022	-0.047**	0.023	2,039

Notes: Parameter $\hat{\beta}$ is the estimated marginal effect from a linear probability model. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

3.2 Heterogeneous Treatment Effects

By Gender. Table 3 examines heterogeneity in recommendation adoption rates by gender. Consistent with our main result, the adoption rate gap is the highest in the scenario with the most objective recommendation basis for either gender. We observe aversion to algorithms across all scenarios, with the exception of female students in the most subjective scenario (*Heart*), who show a marginal, not statistically significant preference for algorithmic recommenders over human counselors, with a 0.8 percentage-point gap ($\hat{\beta}$=0.008, SE=0.032). In the two more objective scenarios (*Geography* and *Pragmatism*), female students adopt recommendations from human counselors at a higher rate than algorithms, with a larger adoption rate gap in the most objective scenario, i.e., we observe a significant 4.9 percentage-point gap in *Pragmatism* versus a—not statistically significant—3.0 percentage-point gap in *Geography* ($\hat{\beta}$=-0.049, SE=0.021; $\hat{\beta}$=-0.030, SE=0.030). For male students, the largest adoption rate gap of 6.5 percentage points is observed in the *Pragmatism* scenario but it is not statistically significant ($\hat{\beta}$=-0.065, SE=0.040).

Given the same recommendation source (algorithmic recommender or human counselor), the magnitude of the adoption rate also differs by gender. For the two more subjective scenarios (*Heart* and *Geography*), female students show consistently lower adoption rate than male students, for both human counselors and algorithmic recommenders. Specifically, there is a 3.4 percentage-point gap ($p = 0.267$) in the descriptive adoption rates of algorithmic recommendations between male and female students in the *Heart* scenario; this gap doubles to 8.1 percentage-points ($p = 0.008$) in the *Geography* scenario. This descriptive gender gap in recommendation adoption *reverses* in *Pragmatism*, the scenario with the most objective recommendation basis: female students adopt recommedations from either source at a much higher rate than male students. The gender gap is 12.2 percentage points ($p < 0.001$) for algorithmic recommenders and 11.0 percentage points ($p < 0.001$) for human counselors. Despite this, in the *Pragmatism* scenario, we also observe the highest level of algorithm aversion for female, as mentioned above.[4]

By Tertile of Prior-year GPA. Table 3 examines heterogeneity in adoption rates by tertiles of prior-year

[4]Table S4 compares the recommendation adoption rate gap between female and male students.

Table 3: Estimated Recommendation Adoption Rates, by Gender and by GPA

	Means		Without Controls		With Controls		
Scenario	Human	Algorithm	$\hat{\beta}$	SE	$\hat{\beta}$	SE	N
By Gender							
Heart							
Female	0.294	0.304	0.005	0.034	0.008	0.032	1,246
Male	0.361	0.338	-0.030	0.022	-0.039*	0.022	747
Geography							
Female	0.339	0.314	-0.032	0.032	-0.030	0.030	1,249
Male	0.417	0.395	-0.030	0.030	-0.027	0.028	752
Pragmatism							
Female	0.514	0.462	-0.047**	0.020	-0.049**	0.021	1,242
Male	0.404	0.340	-0.065*	0.039	-0.065	0.040	742
By Tertile of Prior-year GPA							
Heart							
Bottom	0.344	0.368	0.015	0.051	0.008	0.050	674
Middle	0.289	0.298	0.006	0.029	-0.005	0.028	668
Top	0.317	0.280	-0.039	0.028	-0.029	0.028	678
Geography							
Bottom	0.424	0.405	-0.027	0.019	-0.022	0.020	683
Middle	0.309	0.312	-0.003	0.036	-0.004	0.037	666
Top	0.369	0.308	-0.067***	0.022	-0.069***	0.023	680
Pragmatism							
Bottom	0.420	0.336	-0.082**	0.034	-0.080**	0.033	674
Middle	0.494	0.430	-0.064*	0.036	-0.059	0.038	664
Top	0.500	0.492	-0.004	0.036	-0.006	0.037	673

Notes: Parameter $\hat{\beta}$ is the estimated marginal effect from a linear probability model. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

GPA.[5] In the *Geography* scenario, participants in the top GPA tertile adopt algorithmic recommendations at significantly lower rates compared to human recommendations, with a 6.9 percentage-point gap ($\hat{\beta}$=-0.069, SE=0.023). In the *Pragmatism* scenario, participants in the bottom GPA tertile exhibit a similar preference for human recommendations, with a significant 8.0 percentage-point gap ($\hat{\beta}$=-0.080, SE=0.033). No significant differences are observed in the *Heart* scenario across GPA tertiles. These results indicate a potential mismatch in the preferences of high- and low-achieving students. High-achieving students are more likely to prefer human recommendations in scenarios emphasizing geographic considerations, but show no preference in pragmatic decision-making. In contrast, low-achieving students significantly prefer

[5]Figure S1 shows the histogram of prior-year student performance across all participants.

human counselors in pragmatic decision-making.[6]

By Norm Compliance Levels. Table 4 reports adoption rates by norm compliance tertiles.[7] Participants with high norm compliance scores show a stronger preference for human recommendations in the *Geography* scenario, with a significant 8.3 percentage-point gap ($\hat{\beta}$=-0.083, SE=0.026). Similarly, in the *Pragmatism* scenario, those in the top and middle tertile of norm compliance adopt human recommendations at significantly higher rates ($\hat{\beta}$=-0.066, SE=0.030; $\hat{\beta}$=-0.063, SE=0.033). These findings suggest that norm compliance influences how participants perceive the reliability of recommendation sources, particularly in scenarios where objectivity is central to the recommendation. This result aligns with prior research showing that algorithmic recommendations are more likely to be adopted when they align with social norms (Starke et al., 2021).[8]

By Urbanicity. Table 4 examines differences in adoption rates between rural and urban participants. In the *Pragmatism* scenario, rural participants significantly favor human over algorithmic recommendations, with a 9.3 percentage-point gap ($\hat{\beta}$=-0.093, SE=0.042). In contrast, urban participants show no significant preference between recommendation types in any scenario. These results indicate that rural participants are more sensitive to the source of recommendations in the objective scenario.[9]

By Intent to Participate in National Exams. In Table 4, we explore heterogeneity by participants' intent to take national exams. Focusing on students who intend to participate, we observe a trend similar to the main results, that is, algorithm aversion is the largest and statistically significant in the scenario with the most objective recommendation basis (*Pragmatism*) with an adoption rate gap of 5.0 percentage points ($\hat{\beta}$=-0.050, SE=0.021). In the *Geography* scenario, participants not intending to take the exams exhibit a notably high preference for human recommendations over algorithmic ones, with a significant 12.1 percentage-point difference ($\hat{\beta}$=-0.122, SE=0.037).[10]

Our investigation of heterogeneous algorithm aversion sheds light on the underlying mechanisms. Concerns about algorithmic bias against personal characteristics, such as gender, study preferences, and social compliance, may increase aversion to algorithmic recommendations.[11] Section 5 leverages this

[6]Table S5 compares the recommendation adoption rate gap between tertiles of prior-year GPA.

[7]Figure S2 illustrates the distribution of norm compliance scores across all participants.

[8]Table S6 compares the recommendation adoption rate gap between tertiles of the norm compliance score.

[9]Table S7 compares the recommendation adoption rate gap between students in rural and urban schools.

[10]Table S8 compares the recommendation adoption rate gap between students who intend to participate in the national exams for university admission and those who do not.

[11]Tables S9 and S10 present results from additional heterogeneity analyses based on students' reported intention to use the recommender and their grade level, respectively. We find that students who would not use an algorithmic recommender (if available) exhibit a high level of aversion to algorithmic recommendations in the scenario with the most subjective recommendation basis (*Heart*). Conversely, students who would use the algorithmic recommender show greater aversion in the scenario with the most objective recommendation basis (*Pragmatism*). Moreover, we observe substantial and statistically significant aversion to algorithmic recommenders among grade 10 students in the *Geography* scenario and grade 11 students in the *Pragmatism* scenario. Grade 10 students exhibit statistically comparable recommendation adoption rates between human and algorithmic recommenders across all scenarios. In the *Pragmatism* scenario, the recommendation adoption rate gap for grade 12 students is sizeable but statistically imprecise. Table S12 reports differential adoption rate gaps by tertile of prior-year performance, defined as the average reported scores in mathematics and Greek language. The findings align with the results from the heterogeneity analysis by tertile of prior-year GPA. Table S11 reports differential adoption rate gaps by student track choice. Table S13 compares the recommendation adoption rate gap between students who intend to use the depicted recommender and those who do not. Table S14 examines this gap across different grade levels. Table S15 presents

heterogeneity to propose a framework for policymakers to design effective counseling initiatives.

3.3 Reasons for Resistance to Algorithmic Recommendations

Table 5 presents results from a mediation analysis (Baron and Kenny, 1986),[12] examining how four mediating channels—perceived *intent*, *ability*, *comprehension*, and *alignment*—affect the recommendation adoption rates in the *Heart*, *Geography*, and *Pragmatism* scenarios. These mediators represent key dimensions influencing students' evaluation of recommendations.[13]

We examine how algorithmic recommender assignment influences recommendation adoption through identified mediators. We separate the total effect of the algorithmic recommender assignment on recommendation adoption into indirect effects mediated by each of the four channels and the remaining direct effect. Indirect effect estimates capture the impact of assignment on adoption via each mediator, while direct effect estimates reflect the unmediated effect of assignment on adoption. For example, if recommendations are assigned either by an algorithm or a human counselor, perceived ability may mediate adoption—users might be more likely to follow recommendations from the source they deem more knowledgeable. If human counselors are perceived as having greater expertise in understanding users' needs, their recommendations may be adopted at higher rates, whereas algorithmic recommendations may face skepticism. In this case, the indirect effect captures the extent to which assignment influences adoption through perceived ability, while the direct effect reflects any remaining impact beyond this mechanism.

Perceived recommender *intent* and *ability* are statistically significant mediators of the difference in recommendation adoption rate between human counselors and algorithmic recommenders across all three scenarios. The indirect effect of perceived *intent* ranges from -0.017 (*Heart* scenario) to -0.034 (*Geography* and *Pragmatism* scenarios), leaving small and statistically no significant direct effects of the algorithmic recommender on recommendation adoption. The indirect effect of perceived *ability* ranges from -0.030 (*Pragmatism* scenario) to -0.056 (*Heart* scenario). The direct effects of the algorithmic recommender on recommendation adoption are not statistically significant except for the *Heart* scenario. The positive and significant effect in the *Heart* scenario suggests that when students do not perceive algorithms as having lower ability, they may actually prefer algorithmic recommendations over those from human counselors. In every case, the magnitude of the indirect effect exceeds the magnitude of the direct effect. These results suggest that students who reject algorithmic recommendations do so because they are suspicious of the algorithm's good will and have little faith in how knowledgeable it is. Put differently, there may be little intrinsic aversion to algorithmic recommenders beyond these identified mechanisms.

Reported *comprehension* of the recommender's rationale is found to be a significant mediator of the recommendation adoption rate gap between humans and algorithms only in the two most subjective scenarios of *Heart* and *Geography*. Specifically, the estimated indirect effect in the two scenarios is around 1 percentage point, with the direct effect being almost four times larger and statistically significant

the recommendation adoption rate gap by tertile of prior-year performance, measured as the average score in Mathematics and Greek Language. Table S16 presents the recommendation adoption rate gap by track choice.

[12] We provide an explainer of Mediation Analysis in Appendix Section S9.

[13] Table S17 presents the estimated difference in mediator scores between students in the *algorithmic recommender* and *human counselor* conditions. Students in the *algorithmic recommender* condition have statistically significantly lower scores in the mediators *intent* and *ability* compared to those in the *human counselor* condition across all scenarios. The corresponding difference in *comprehension* is positive and statistically significant only in the *Geography* scenario, while the difference in the *alignment* mediator is negative and statistically significant only in the *Pragmatism* scenario.

in the *Geography* scenario. The estimated direct effect associated with the algorithmic recommender after accounting for perceived *comprehension* is not statistically significant in the *Heart* scenario. In the *Pragmatism* scenario, the indirect effect of *comprehension* is small and statistically not significant, leaving a large, negative, and significant direct effect on recommendation adoption associated with the algorithmic recommender. These results suggest that in the *Geography* and *Heart* scenarios, where the recommendation basis is more subjective, the algorithmic recommender is associated with higher reported *comprehension* (see Table S17). This suggests that the algorithmic recommender may appear less nuanced compared to human counselors when it gives recommendations that are not as objective.

Perceived *alignment* of the recommendation with the personal preferences does not significantly mediate the adoption rate gap between human counselors and algorithmic recommenders in the *Geography* and *Heart* scenarios. However, *alignment* is found to significantly mediate algorithmic aversion in the *Pragmatism* scenario. In other words, students who reject the algorithmic recommendation in the *Pragmatism* scenario may do so because they consider it as invalidating of personal preferences. After accounting for the indirect effect of perceived *alignment* on recommendation adoption, the algorithmic recommemder is associated with a negative direct effect of equal magnitude, albeit statistically not significant.

Each mediator captures distinct yet somewhat overlapping dimensions of students' decision-making processes. Table S22 reports Cronbach's alpha coefficients for different combinations of mediators across the three decision-making scenarios.[14] The results show that combining more mediators generally increases reliability, with the highest alpha values observed when all four mediators are included. Individual mediator pairs, such as *comprehension* and *alignment*, exhibit relatively low reliability, while broader combinations yield stronger consistency across scenarios. Importantly, none of the pairs yield extremely high alpha values, underscoring the robustness of the mediators in capturing a wide range of underlying factors influencing students' decision-making processes.

We also investigate the collective mediating influence of all four mediators. Table S18 presents the results of this collective mediation analysis, examining how specific mediators contribute to the observed adoption rate gap between human counselors and algorithmic recommenders across the three scenarios.[15] The results reveal that all four mediators significantly explain part of the effect of the algorithmic recommender, albeit with varying degrees of contribution depending on the scenario. In the *Geography* scenario, *intent* and *alignment* emerge as the two largest significant mediators, with indirect effects of 0.036 and 0.047 (both $p < 0.01$), respectively. Similarly, *comprehension* plays a notable role in the *Pragmatism* scenario, with an indirect effect of 0.030 ($p < 0.01$). Notably, the *Heart* scenario highlights *intent* as the strongest mediator, with an indirect effect of 0.021 ($p < 0.01$). At the same time, the direct effect of the algorithmic recommender in magnitude and loses statistical significance when all mediators are included, indicating that the algorithmic recommender's influence operates primarily through the identified mediators rather than independently.

Overall, our results highlight that all four mediators influence recommendation adoption across dif-

[14] Cronbach's alpha is a measure of internal consistency, indicating how well a set of variables in a test or survey measures a single latent construct. A low Cronbach's alpha can indicate that the questions in a test measure different aspects of a topic rather than a single, unified idea, i.e., capture multiple dimensions or perspectives.

[15] Tables S19-S21 present the mediating effects of different mediator combinations on the impact of the algorithmic recommender on recommendation adoption in each scenario. Each model iteratively excludes one mediator to assess the sensitivity of the remaining mediator coefficients.

ferent scenarios, albeit to varying extents. However, consistently across all scenarios and various statistical tests presented in Table 5 and Appendix Section S4, *intent* stands out as a key mediator. Its indirect effect on recommendation adoption is significant across all scenarios (Table 5) and its effect remains significant when we control for the other mediators in all scenarios (Appendix Section S4). Focusing on the scenario with the highest algorithmic aversion (*Pragmatism*), *alignment* also proves to be an important and statistically significant channel, with its effect on recommendation adoption being the largest when we control for the other mediators.

Table 4: Estimated Recommendation Adoption Rates, by Norm Compliance Levels, by School Urbanicity, and by Intention to Participate in National Exams

	Means		Without Controls		With Controls		
Scenario	Human	Algorithm	$\hat{\beta}$	SE	$\hat{\beta}$	SE	N
By Tertile of Norm Compliance Score							
Heart							
Bottom	0.229	0.241	0.011	0.038	0.005	0.037	630
Middle	0.305	0.296	-0.018	0.039	-0.019	0.039	585
Top	0.403	0.397	-0.015	0.043	-0.015	0.043	788
Geography							
Bottom	0.339	0.313	-0.031	0.043	-0.034	0.041	631
Middle	0.315	0.359	0.034	0.030	0.029	0.027	590
Top	0.421	0.340	-0.090***	0.028	-0.083***	0.026	788
Pragmatism							
Bottom	0.498	0.487	-0.014	0.033	-0.017	0.037	633
Middle	0.487	0.406	-0.073**	0.031	-0.066**	0.030	590
Top	0.439	0.372	-0.064**	0.033	-0.063*	0.033	791
Rural vs. Urban							
Heart							
Rural	0.322	0.395	0.051	0.056	0.045	0.053	593
Urban	0.313	0.288	-0.024	0.024	-0.024	0.022	1,455
Geography							
Rural	0.362	0.367	-0.032	0.038	-0.028	0.043	596
Urban	0.367	0.334	-0.031*	0.017	-0.030*	0.017	1,461
Pragmatism							
Rural	0.454	0.350	-0.089**	0.038	-0.093**	0.042	592
Urban	0.479	0.445	-0.032	0.024	-0.033	0.023	1,447
By Intent to Participate in National Exams							
Heart							
Yes	0.315	0.314	-0.004	0.022	-0.009	0.021	1,719
No/Other	0.316	0.335	0.007	0.058	0.012	0.056	329
Geography							
Yes	0.339	0.335	-0.010	0.019	-0.012	0.018	1,722
No/Other	0.494	0.384	-0.125***	0.035	-0.122***	0.037	335
Pragmatism							
Yes	0.491	0.435	-0.054***	0.021	-0.050**	0.021	1,710
No/Other	0.372	0.331	-0.038	0.072	-0.041	0.069	329

Notes: Parameter $\hat{\beta}$ is the estimated marginal effect from a linear probability model. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 5: Mediation Analysis

	Intent		Ability		Comprehension		Alignment	
Scenario	$\hat{\beta}$	SE	$\hat{\beta}$	SE	$\hat{\beta}$	SE	$\hat{\beta}$	SE
Heart								
Indirect Effect	-0.017***	0.006	-0.056***	0.009	0.010**	0.005	0.012	0.009
Direct Effect	0.010	0.020	0.047**	0.024	-0.018	0.022	-0.020	0.016
Mediator Mean	1.217		1.066		1.704		-1.500	
Mediator SD	2.851		2.735		2.791		3.190	
Geography								
Indirect Effect	-0.034***	0.007	-0.038***	0.014	0.011***	0.004	-0.001	0.007
Direct Effect	0.004	0.019	0.009	0.020	-0.042**	0.017	-0.029**	0.014
Mediator Mean	1.392		1.328		2.469		0.386	
Mediator SD	2.365		2.620		2.378		2.918	
Pragmatism								
Indirect Effect	-0.034***	0.008	-0.030***	0.008	-0.002	0.004	-0.024***	0.007
Direct Effect	-0.013	0.020	-0.017	0.023	-0.045**	0.023	-0.024	0.024
Mediator Mean	2.047		1.637		2.290		1.320	
Mediator SD	2.387		2.503		2.547		2.845	

Notes: Participants rated each mediating channel's associated statement on a scale ranging from -5 to 5. $\hat{\beta}$ is the estimated effect from a linear probability model. Regression controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

4 Insights from Students' Free-text Responses

In addition to the quantitative analysis presented in Section 3.3, we conducted a qualitative analysis of students' free-text responses to the question, *"How could such a counselor [as in the recommendation source] be more helpful to you?"* Students were asked to provide their input with respect to the specific recommendation source that was assigned to them, that is, either a human counselor or an algorithmic recommender. The framework provides a structured approach to understanding how students value and interpret human and algorithmic guidance in their educational decisions.

The responses, originally provided by the students in Greek, were translated into English using DeepL Translate, with additional manual quality checks performed by the research team. These responses were then thematically categorized into three key themes: **Helpfulness**, **Purpose**, and **Recommendation Basis**. Under the topic of **Helpfulness**, responses were categorized as *Helpful*, *Not Helpful*, or *Unknown*, based on how students perceived the value of the recommender. The **Purpose** theme captured students' intended use of the recommender, distinguishing between *Decide for Me* (deferring the decision to the recommender), *Direction* (seeking guidance), *Exploration* (considering new options), and *Validation* (confirming intended choices). Finally, the **Recommendation Basis** theme reflected the factors students considered important in the recommendation process, such as *Career Prospects*, *Grades and Chances*, and *Interests*. Select responses and their corresponding classifications are listed in Table 6.

Table 6: Select Text Responses

Topic	Text Responses
Helpfulness	
Helpful	[H] *"The counselor would help me with more information."*
Unknown	[A] *"I have no idea."*
Not Helpful	[A] *"I wouldn't trust an algorithm for such an important decision. No matter how much it improved, I wouldn't use it."*
Purpose	
Decide for Me	[H] *"To direct me to the school that he thinks suits me based on his experience."*
Direction	[H] *"To guide me according to my interests and performance."*
Exploration	[A] *"It could be useful for me to see what suits me and to see what other options I have."*
Validation	[A] *"It would confirm if the profession I have chosen to study is suitable for me."*
Recommendation Basis	
Career Prospects	[H] *"Telling me which school has the best professors and programs, but also the weight or value of the school's degree in finding a job."*
Grades and Chances	[A] *"Make suggestions for schools that I can get into depending on my grades."*
Interests	[H] *"Suggest something that fits their character and build on what the student likes."*

Notes: The table lists select student responses for a free-text question, *"How could such a counselor [as in the recommendation source] be more helpful to you?"*, and the corresponding classification. [A] and [H] denote the recommendation sources, where [A] represents the algorithmic recommender and [H] represents the human counselor for the student response.

We conducted zero-shot classification using a pre-trained large language model to determine the classification scores of student responses for each topic-category combination. We employed the DeBERTa-v3 transformer model, which is pre-trained and fine-tuned for natural language inference, to independently classify responses and score the topic categories (Sileo, 2024). These scores reflect the probability of the response belonging to the respective topic-category combination. The approach enabled classification without the need for additional training or labeled data, ensuring both consistency and adaptability across our diverse topics. The analysis focused on the text responses of students who, in a prior survey question, expressed a willingness to receive personalized recommendations from either a human or algorithmic counselor. The question asked, *"I would like to receive personalized recommendations on my college program choices from a counselor [as in the recommendation source] like the one in the stories,"* with students scoring it on a Likert-type scale from -5 to +5. Only responses from students who scored zero or above on this question were included. These students are referred to as *usage-inclined*. Out of the 1,823 text responses, 1,402 (77%) came from *usage-inclined* students.

Table 7: Text Classification Scores

Topic	Human Counselor	Algorithmic Recommender	Difference
Helpfulness			
Helpful	0.78 (0.76-0.80)	0.67 (0.65-0.70)	-0.105***
Unknown	0.12 (0.11-0.13)	0.16 (0.14-0.18)	0.040***
Not Helpful	0.01 (0.01-0.01)	0.02 (0.01-0.03)	0.010***
Purpose			
Decide for Me	0.15 (0.14-0.17)	0.16 (0.14-0.17)	0.005
Direction	0.34 (0.32-0.36)	0.27 (0.25-0.29)	-0.070***
Exploration	0.28 (0.27-0.30)	0.30 (0.28-0.32)	0.014
Validation	0.30 (0.29-0.32)	0.30 (0.28-0.31)	-0.006
Recommendation Basis			
Career Prospects	0.44 (0.42-0.46)	0.37 (0.35-0.40)	-0.064***
Grades and Chances	0.29 (0.28-0.31)	0.33 (0.31-0.35)	0.040***
Interests	0.52 (0.50-0.54)	0.47 (0.44-0.49)	-0.053***

Notes: This table shows the mean classification scores (with 95% confidence intervals in parentheses) of usage-inclined students' response to the question, *"How could such a counselor [as in the recommendation source] be more helpful to you?"*, across three topics. Usage-inclined students are those who scored zero or higher on a Likert-type scale ranging from -5 to +5, indicating their willingness to use a recommender for college application decisions. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 7 shows the classification scores for student responses across the three topics. Consistent with the analysis in Section 3, the results indicate that students are more likely to find human counselors more helpful overall, compared to algorithmic recommenders. They are also more likely to find human counselors more effective in providing direction for decision-making. In terms of the recommendation basis, students are slightly more inclined to find algorithmic recommenders helpful for determining grade requirements and their chances of getting into college programs. However, when the recommendation requires a more subjective or highly personalized assessment of personal interests or career prospects, students considered human counselors more helpful.[16] Our results align with findings from a study on learning analytics tools in Finland, where users valued the ability to explore career options and decision support but sought further personalization and direction on career paths (Gedrimiene et al., 2023).[17] We also conducted a sentiment analysis, with responses categorized as *Positive*, *Negative*, and *Neutral*. Our results, illustrated in Table S24, are in line with the primary analysis in Table 7.[18]

5 Policy Application: Targeted Treatment Assignment

The findings in Section 3.2 highlight that there is significant heterogeneity in adoption rates among different student subgroups. In an ideal setting with no cost constraints, a social planner seeking to maximize recommendation adoption would allocate human counselors and algorithmic recommenders on an individual basis, selecting the most effective intervention for each student's characteristics. However, the cost and limited availability of human counselors, coupled with varying levels of students' resistance to (or preference for) algorithmic recommenders, create an inherent trade-off: Should the available human counselors be allocated to students who exhibit the greatest aversion to algorithmic recommenders, or should they be directed toward those more receptive to recommendations, regardless of the source? Ultimately, we show that the optimal allocation critically depends on the planner's priorities. Since a tailored approach to this allocation requires collecting potentially costly information about each student, we also ask: Can targeting policies that rely on coarser, less personalized criteria be sufficiently effective?

In Section 5.1, we begin by estimating the distribution of heterogeneous treatment effects within the sample. In Section 5.2, we formulate an optimization problem where the objective reflects the planner's varying prioritization of adoption rates. In Section 5.3, we identify the first-best targeting policy and compare it to simple heuristics that rely on small sets of student characteristics, quantifying the resulting loss. Finally, we examine optimal targeting under different planner priorities.

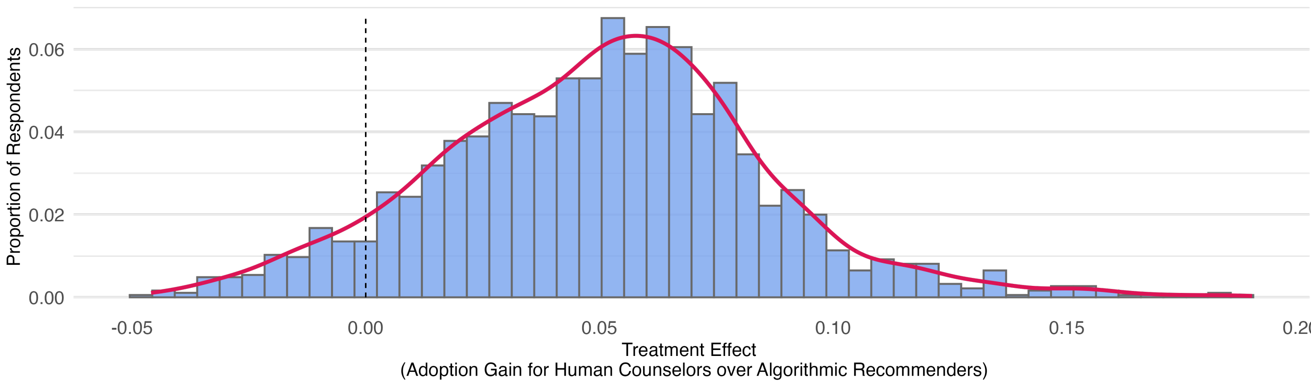
[16] We conducted text classification analysis of free-text responses by student characteristics (see Tables S30 to S35).

[17] Tables S25-S27 report the correlation coefficients between text topic classification scores and our mediators in each scenario. While the identified text topics are associated with our mediators—particularly in the *Pragmatism* scenario—Table S28 indicates that their mediating influence on algorithmic aversion is weak. This suggests that the preference for human versus algorithmic recommendations may depend more on scenario-specific contextual details than on student-level beliefs.

[18] For completeness, we also conducted the text classification analysis for the free-text responses of students who, in a prior survey question, expressed a willingness to *not* receive recommendations from either a human counselor or an algorithmic recommender. This comprised 23% (421) of all respondents. The results, presented in Table S29, show no significant differences across topic categories except for *helpful* between human counselors and algorithmic recommenders for students not inclined to use recommenders.

5.1 Estimating CATEs using a Causal Forest

Figure 1: HETEROGENEITY IN TREATMENT EFFECT



Notes: This figure shows the frequency distribution of treatment effect estimates derived from the Causal Forest for the *Pragmatism* scenario. Figures S6 and S7 plot the corresponding treatment effect distribution in the *Heart* and *Geography* scenarios, respectively. The spread of the histogram reflects the heterogeneity in treatment effects, indicating variations in adoption gains when a human counselor is provided compared to an algorithmic recommender. We illustrate the heterogeneity of CATE across gender in Figure S4, and find that female students tend to experience higher treatment effects on average compared to male students, but the spread of the effect is higher among male students.

To better understand the heterogeneity in intervention efficacy, we adopt a Causal Forest[19] model to estimate *Conditional Average Treatment Effects* (CATEs). CATE measures the expected treatment effect for a specific subgroup of participants with similar features; the Causal Forest model is a machine-learning method that flexibly estimates these effects by leveraging decision trees to identify patterns of heterogeneity (Athey and Wager, 2019, 2021). In this context, we define algorithmic recommenders (A) as the control group and human counselors (H) as the treatment group. Figure 1 illustrates the distribution of treatment effects across the whole population, indicating the presence of potential heterogeneity in treatment effects.[20] In Figure S3, we assess the relative importance of participant characteristics in predicting treatment effects and observe significant variation across different characteristics. We find that GPA and norm compliance have the highest importance scores.

5.2 Setting up the Optimization Problem of Treatment Assignment

The optimization problem aims to assign human counselors or algorithmic recommenders to student subgroups to maximize the planner's objective, subject to the limited number of available human counselors. The process involves three main steps. First, student subgroups are defined based on relevant features

[19] Causal Forests, an extension of the Random Forest algorithm, are specifically designed to estimate the causal effect of a treatment on an outcome variable. Unlike traditional Random Forests, which focus on predictive accuracy, Causal Forests aim to uncover the heterogeneity in treatment effects by maximizing the difference in treatment effects across various data splits within the tree structure. This approach allows for the identification of population subgroups within the data where treatment effects may differ. We provide an explainer in Appendix S10.

[20] We fit the Causal Forest on the data from the *Pragmatism* scenario; this scenario aligns with students' intended use of algorithmic recommenders (prediction of admission chances based on grades), as shown in the text analysis in Section 4. Results based on data from the *Heart* and *Geography* scenarios are provided in Appendix S7.

such as gender, GPA, urbanicity, and norm compliance. We test policies that use either finer distinctions to capture granular differences or coarser groupings to reduce model complexity and information collection costs. Second, outcomes for these subgroups are estimated using a Causal Forest model (as in Section 5.1), which predicts treatment effects by leveraging both individual- and group-level characteristics. Third, an optimization model is used to determine the optimal allocation of human counselors, while the remainder of the population is assigned to algorithmic recommenders.

Formally, let $\mathcal{X}$ denote the collection of all subgroups defined in the first step. For a specific group $x \in \mathcal{X}$, the planner's objective is given by:

$$\begin{aligned} f(x) &= A(x)S(x)g(\hat{\theta}_H(x)) + (1 - A(x))S(x)g(\hat{\theta}_A(x)) \\ &= A(x)S(x)\left(g(\hat{\theta}_H(x)) - g(\hat{\theta}_A(x))\right) + S(x)g(\hat{\theta}_A(x)). \end{aligned}$$

$A(x)$ denotes the allocation decision for group x, ranging between 0 and 1, and represents the proportion of the group allocated to human counselors. $S(x)$ is a weighting function, indicating the proportion of individuals in the population with features x. The outcomes (recommendation adoption) are denoted by $\hat{\theta}_H(x)$ and $\hat{\theta}_A(x)$, respectively, representing the expected probabilities of adoption for an individual with features x when assigned to the human counselor(H or Algorithmic recommender (A. For instance, the group of male students in Exact Sciences at rural schools (comprising $S(x) = 3.8\%$ of the population) have the lowest algorithmic adoption rate ($\hat{\theta}_A = 0.323$), and the highest estimated gain in adoption rate when assigned to human counselors ($\hat{\theta}_H = 0.407$, $\hat{\theta}_H - \hat{\theta}_A = 0.084$).

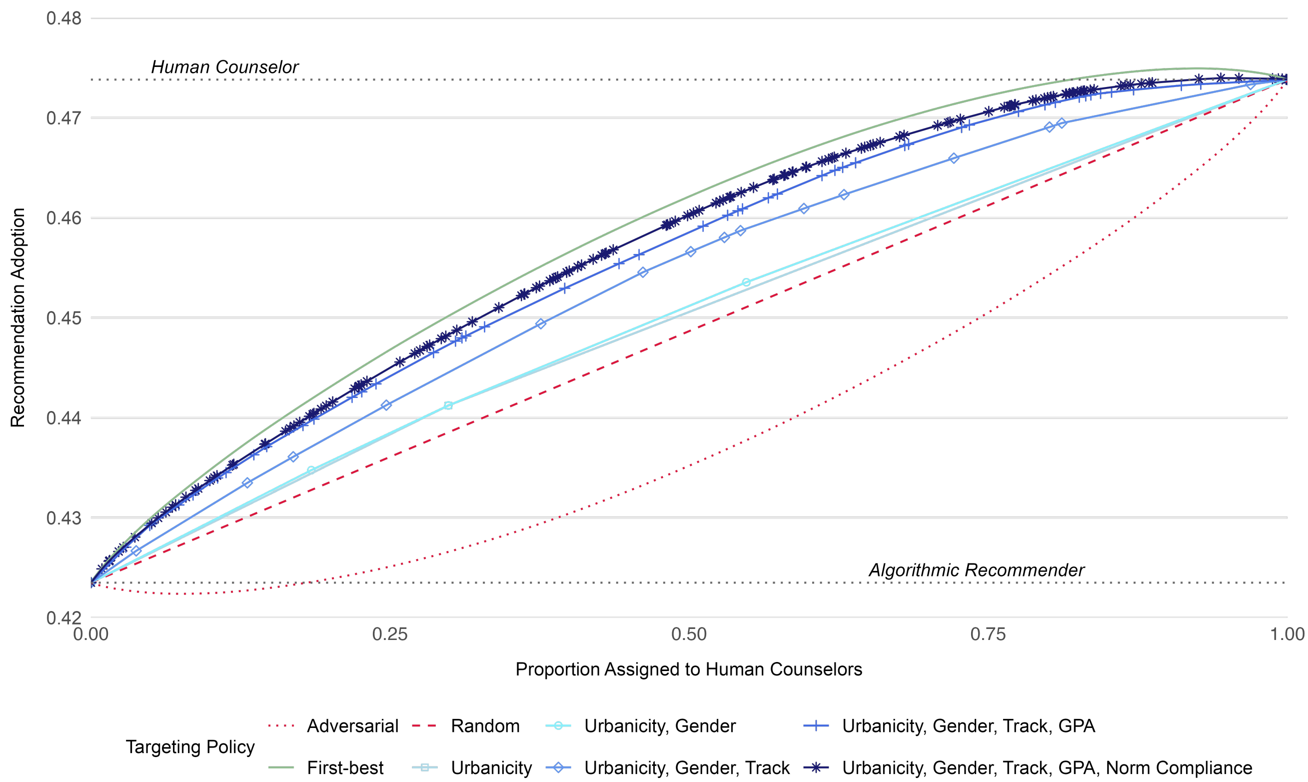
Given the limited capacity of human counselors, the choice of $g(p)$ reflects the planner's prioritization of different control subgroups. The planner must decide whether to (i) maximize adoption rate improvements (i.e., they have an equal preference for adoption rate increments at any baseline level of adoption), or (ii) prioritize adoption rate improvements among individuals with lower baseline adoption rates, or (iii) prioritize adoption rate improvements among individuals with higher baseline adoption rates. Here, the baseline adoption rate refers to the likelihood of adoption under assignment to the algorithmic recommender. The planner prioritization preferences align with the framework proposed by Fernández-Loría and Provost (2022) and are referred to as the planner's focus on *Persuadables, Lost Causes,* and *Sure Things* respectively. The planner's preference is formalized using a function $g(p)$, where $g(p)$ determines the priority assigned based on the baseline adoption probability p. The planner's preference is reflected in the sign of the second derivative: (i) $\frac{\partial^2 g(p)}{\partial p^2} = 0$ implies prioritizing the maximization of adoption rate (focus on *Persuadables*), (ii) $\frac{\partial^2 g(p)}{\partial p^2} < 0$ implies prioritizing individuals with lower baseline adoption (focus on *Lost Causes*), and (iii) $\frac{\partial^2 g(p)}{\partial p^2} > 0$ implies prioritizing individuals with higher baseline adoption (focus on *Sure Things*).

Given a capacity constraint C, representing the proportion of students that can be assigned to a human counselor, the corresponding optimization problem is defined as follows:

$$\begin{aligned} \max_{A(x)\, \forall x \in \mathcal{X}} \quad & A(x)S(x)\left(g(\hat{\theta}_H(x)) - g(\hat{\theta}_A(x))\right) + S(x)g(\hat{\theta}_A(x)) \\ \text{s.t.} \quad & \sum_x A(x)S(x) \leq C, \\ \text{and} \quad & 0 \leq A(x) \leq 1. \end{aligned} \tag{2}$$

5.3 Optimized Treatment Assignment

Figure 2: TARGETING POLICY PERFORMANCE



Notes: This figure shows the overall adoption rate under different targeting policies as the share of the targeted population increases in the *Pragmatism* scenario. Figures S8 and S9 plot the corresponding targeting policy performance in the *Heart* and *Geography* scenarios, respectively. For these results, we assume the planner focuses on *Persuadables* (i.e., $g(p) = p$). The area under each curve (AUC) represents the overall adoption rate when the entire population is targeted. The performance of a targeting policy is measured by the difference between the AUC under a specific policy and the AUC under a baseline policy. (In Table 8, we consider either the adversarial policy or the random policy as the baseline.) The *first-best* targeting policy prioritizes the allocation of human counselors to individuals with the highest estimated treatment effects, representing the best possible allocation policy in terms of efficiency. Conversely, the *adversarial* targeting policy prioritizes the allocation of human counselors to individuals with the lowest estimated treatment effects, representing the worst possible allocation policy in terms of efficiency. Both the first-best and adversarial targeting policies involve individual-level targeting based on estimated individual treatment effects. The *random* targeting policy allocates human counselors uniformly at random, independent of the student's characteristics. The remaining strategies involve subgroup-level targeting, utilizing information about school locale (urbanicity), student gender, track choice, prior-year Grade Point Average (GPA), and norm compliance.

In this section, we present our optimization results for various classes of targeting policies and planner priorities. Figure 2 illustrates the performance of targeting policies that use varying combinations of school and student information to determine the treatment assignment. The recommendation adoption (on the *y-axis*) is presented across different capacity constraints (i.e., the proportion of population assigned to available human counselors). We assume that the planner focuses on *Persuadables* (i.e., $g(p) = p$), with results for other planner priorities shown in Figure S5.

First-best Targeting Policy. Under the first-best targeting policy, the planner would know each individual's likelihood of adopting a recommendation under both a human counselor and an algorithmic recommender, thus allowing for *personalized* treatment assignment. The first-best policy minimizes misallocation at the most granular (individual) level. It assigns human counselors to individuals who experience the highest improvement in recommendation adoption from human counselors compared to algorithmic recommenders. This represents the best outcome achievable by a policymaker who can tailor decisions to each individual based on their adoption behavior. Other targeting policies can be evaluated against this benchmark to assess their relative effectiveness.

Table 8: TARGETING POLICY PERFORMANCE

Targeting Policy	**AUC**	**Performance (%)**
First-best	0.45830	100.0
Urbanicity, Gender, Track, GPA, Norm Compliance	0.45682	92.3
Urbanicity, Gender, Track, GPA	0.45591	87.6
Urbanicity, Gender, Track	0.45375	76.4
Urbanicity, Gender	0.45031	58.6
Urbanicity	0.45000	57.0
Random	0.44865	50.0
Adversarial	0.43901	0.0

Notes: AUC (Area Under the Curve) represents the overall adoption rate under a counterfactual targeting policy, such as adversarial or random allocation. Targeting policy performance is calculated as the difference between the AUC under a specific policy and the adversarial policy relative to the performance of the first-best policy over the adversarial policy. For these results, we assume that the planner focuses on *Persuadables* (i.e., $g(p) = p$). The first-best targeting policy prioritizes the allocation of human counselors to individuals with the highest estimated treatment effects, representing the best possible allocation policy. Conversely, the adversarial targeting policy prioritizes the allocation of human counselors to individuals with the least estimated treatment effects, representing the worst possible allocation policy. Both the first-best and adversarial targeting policies assume individual-level targeting based on estimated individual treatment effects. The remaining strategies assume subgroup-level targeting, using information on school locale (urbanicity), student gender, track choice, prior Grade Point Average (GPA), and norm compliance.

Adversarial and Random Targeting Policies. We calculate the adoption rate of the first-best policy relative to the adversarial policy, which prioritizes assigning human counselors to the individuals with least improvement in adoption. Assigning human counselors at random would generate 50% of the first-best policy's adoption rate gains relative to the adversarial policy.

Targeting Policies with Varying Levels of Granularity. Policymakers must balance the potential gains in adoption from more nuanced targeting strategies against the added complexity and cost. For example, targeting based solely on urbanicity is more straightforward than using urbanicity, gender, GPA, track choice, and norm compliance. Similarly, collecting additional information on personality and psychological measures, such as norm compliance, is costly. We explore how targeting policies using subsets of urbanicity, gender, track, GPA, and norm compliance perform.

We find that using five school and student characteristics—school locale (urbanicity), student gender, track choice (e.g., humanities or science), prior-year Grade Point Average (GPA), and a student's norm compliance in college applications—produces 92.3% of the adoption rate gains achieved by the first-best targeting policy relative to the adversarial policy (Table 8). Further, we observe that, relying solely on data typically found in administrative records—such as school locale (urbanicity), student gender, track choice (e.g., humanities or science), and prior-year Grade Point Average (GPA)—still achieves 87.6% of the first-best policy's adoption gains. We view these gains as satisfactory for a first-round policy aimed at achieving meaningful impacts with manageable costs. The effectiveness of the targeting approach influences how many students a planner needs to assign to human counselors in order to achieve a desired adoption rate. For example, to achieve a 45% adoption rate, targeting with four characteristics performs nearly as well as first-best (personalized) targeting, requiring only 33% of students assigned to human counselors, compared to 30% with the first-best targeting policy. In contrast, relying on just two characteristics is drastically less effective, requiring 54% of students to be assigned to human counselors.

Targeting under Different Planner Priorities. Planner priorities shape resource allocation, with different objectives driving distinct counselor allocations across demographics. Yet, certain subgroups consistently emerge as critical, as seen when focusing on *Lost Causes* versus *Persuadables.* In Table 9, we illustrate the differences in targeted subgroups under different planner priorities. For example, with a focus on *Sure Things* ($g(p) = p^2$), which places greater emphasis on assigning human counselors to individuals with higher baseline (algorithmic) adoption, the planner would first allocate counselors to *female* students in the *Life Sciences* track at *rural* schools. However, if the planner focuses on *Lost Causes*, i.e., prioritizes groups with lower baseline adoption using $g(p) = p^{0.5}$, they would assign human counselors first to *male* students in the *Exact Sciences* track at *rural* schools. The results for the various planner preference scenarios, when the prior-year GPA information is included for targeting, are presented in Tables S36, S37, and S38.[21] We present the results for different targeting policies under different planner priorities in Figure S5. We find similar trends of marginal reduction in adoption gains as we reduce the number of characteristics used for targeting across all three planner priorities.

[21] Table S39 presents the results from an alternative approach that uses a decision tree to identify target student subgroups based on estimated outcomes from a causal forest, incorporating all available school and student attributes to optimize human counselor assignment. The results are similar to the primary analysis, though the number of characteristics to define each subgroup varies.

Table 9: Subgroup Prioritization for Human Counselor Assignment

Order	Urbanicity	Gender	Track	Share	Adoption Rate
Focus on *Persuadables*					
1	Rural	Male	Exact Sciences	0.0378	0.4267
2	Urban	Male	Exact Sciences	0.1307	0.4335
3	Rural	Female	Life Sciences	0.1691	0.4361
4	Rural	Female	Humanities	0.2469	0.4413
Focus on *Lost Causes*					
1	Rural	Male	Exact Sciences	0.0378	0.4267
2	Urban	Male	Exact Sciences	0.1307	0.4335
3	Rural	Female	Humanities	0.2085	0.4386
Focus on *Sure Things*					
1	Rural	Female	Life Sciences	0.0384	0.4261
2	Rural	Female	Humanities	0.1162	0.4313
3	Urban	Male	Exact Sciences	0.2091	0.4381

Notes: The table displays the top subgroups that the planner should prioritize for human counselor assignment across three priority scenarios. The baseline (algorithmic) adoption is 0.4235. In the first scenario (*Persuadables*), the planner aims to maximize adoption rate improvements by treating all baseline adoption levels equally in terms of preference for adoption rate increments ($g(p) = p$). In the second scenario (*Lost Causes*), planners prioritize adoption rate improvements among individuals with lower baseline adoption rates ($g(p) = p^{0.5}$). In the third scenario (*Sure Things*), planners prioritize adoption rate improvements among individuals with higher baseline adoption rates ($g(p) = p^2$). We assume that the planner has access to information regarding school locale, student gender, and track choice.

6 Discussion and Implications for Policy and Algorithm Design

Our main results in Section 3 show that human counselors outperform algorithmic recommenders in recommendation adoption when the criteria are least subjective, even though some student subgroups exhibit a clear preference for algorithmic recommendations (see Figures 1, S4, S6, and S7). Algorithm aversion is strongest when the recommendation basis is the most objective (*Pragmatism*). An objective recommendation is one grounded in quantifiable and measurable facts with minimal ambiguity (Castelo et al., 2019). Conversely, algorithmic recommenders achieve comparable effectiveness only when the recommendation basis is the most subjective. This pattern underscores a limitation of algorithms in studies or career counseling: delivering objective and potentially grounding recommendations may require incorporating human-like qualities such as empathy, social understanding, ability to accommodate exceptions—qualities that algorithms are perceived to lack (Lee, 2018).

Our mediation analysis further supports this argument. Perceived *intent* of the recommendation source is the most consistent driver of the recommendation adoption gap. As recommendation basis shifts from being the least to the most objective, the mediating influence of perceived *ability* on algorithm aversion narrows. At the same time, the mediating influences of *intent* and *alignment* increase (see Table 5). Put differently, increasing the objectivity of the recommendation basis erodes trust in algorithmic recommenders' intentions and alignment with individual goals compared to human counselors (see Table S17 and 5). Interestingly, *alignment* plays a stronger mediating role than *intent* in the most objective scenario (*Pragmatism*) where the recommender optimizes for admissions chances (see Tables S18). One possible hypothesis for future research is that students fear that the algorithm's global optimization objective clashes with their personal goals. While an algorithm may optimize for social welfare (similar to the planner in Section 5), a human counselor, advising fewer students on a personal level, may offer recommendations that feel more tailored to individual needs rather than optimizing for the broader collective. Beyond the above findings, we observe statistically not significant direct effects of algorithmic recommendations on adoption, suggesting that, aside from the identified mechanisms, there may be little intrinsic aversion to algorithmic recommenders.

An important insight from our analysis is that students' adoption rates exhibit substantial heterogeneity across various characteristics, especially gender, urbanicity, prior academic performance, and norm compliance. For example, in the most objective scenario (*Pragmatism*), we find that female students adopt recommendations at a notably higher rate than male students, but continue to exhibit significant algorithm aversion. Similarly, rural students as well as low-achieving students also show significant aversion to algorithms. Even though the *Pragmatism* scenario features an objective recommendation criterion, the actual recommendation still challenges the norm. This helps explain why students with higher compliance to the norm of "prestige-chasing" show a stronger preference for human counselors in this scenario.

Overall, these results highlight that mistrust in algorithms is intertwined with social characteristics. Efforts such as auditing data and algorithms for discrimination, increasing algorithmic transparency, and addressing misconceptions through educational campaigns can help improve algorithm reliability, mitigate self-selection biases, and ensure equitable access to algorithmic recommendations for all students. Furthermore, incorporating human-like attributes, like relatability or compassion, could help deliver norm-breaking recommendations, as they may reduce the resistance associated with such advice. An open question for future research is whether algorithmic recommendations should be tailored to a student's gender or other characteristics, especially if training data contain inherent biases (Goldfarb and Tucker, 2019).

Another takeaway from our policy exercise in Section 5 is the potential of combining algorithmic recommenders with human counselors for two main reasons. First, students' preferences for human counselors vary across student subgroups (see Sections 3.2 and 5.1). For example, male students in the Economics track at rural schools tend to resist recommendations irrespective of the source, while female students in the Exact Sciences track at rural schools have the highest adoption rate of algorithmic recommendations. Second, algorithmic recommenders are more cost-effective and scalable, making them a practical solution for broader implementation when human counselor availability is limited. We show that targeting policies relying solely on data commonly available in administrative records can achieve satisfactory results with respect to the first-best solution at manageable costs. We acknowledge that such hybrid targeting approaches could potentially lead to unintended inequities across student subgroups depending on the planner priorities and availability of human counselors. For example, if the planner had about 25% capacity of human counselors and cared about maximizing recommendation adoption, then the optimal allocation shown in Table 9 would be both efficient and balanced in terms of gender.

Our study also creates research opportunities for exploring different experimental design choices. In our experiment, participants evaluated decision-making scenarios in a third-person context—a common approach in the literature (Sunstein and Gaffe, 2024)—which allowed for more controlled observations and reduced the likelihood of aspirational responses. However, as theorized by Logg and Schlund (2024), willingness to follow algorithmic advice may be greater in real judgments than in hypothetical scenarios. Future research could test whether in real-world, high-stakes decisions, students are more likely to adopt algorithmic advice.

The findings of this study can extend beyond the specific context of Greece. Approximately 40% of countries have centralized systems for higher education, similar to Greece (Neilson, 2024). For example, India, China, Chile, and several European countries utilize national exams and centralized admissions systems. Further, the widespread use of social media and AI has fostered a generation—including Greek adolescents, as in our sample—with high technological familiarity and a predilection for algorithmic systems (AlfaVita, 2024; Kathimerini, 2023, 2024; Tsitsika et al., 2014; Zhu et al., 2024). These factors may amplify the relevance and adaptability of the study's insights to other settings.

Finally, our approach of understanding students' preferences before deploying a recommendation system can serve as a blueprint for developing effective systems for college applications, education, and beyond, as several of our insights may extend to other algorithmic applications. High-stakes decision-making systems, like those used in college admissions, carry several potential risks and must be designed with the utmost attention to ethical considerations, ensuring fairness, transparency, and accountability. Specifically, certain subgroups may ignore the recommendations, or the system might not be used as intended. For example, our text analysis in Section 4 shows that students perceive algorithmic recommenders as better suited for assessing admission chances. Since biases in beliefs about admission chances can lead students to suboptimal decisions (Bobba and Frisancho, 2022; Larroucau et al., 2024), this highlights a promising opportunity for algorithms to help guide students towards more informed decisions. However, misuse or skewed use of such algorithms could exacerbate disparities. For instance, if female and male students adopt algorithmic recommendations at different rates, as shown in our study, any improvements from the recommendation system might inadvertently increase gender disparities in the outcomes. A user-first approach can guide the successful design of AI systems that better align with students' needs and resonate with the target audience, prior to real-world implementation.

7 Conclusion

This paper contributes empirical evidence to theoretical discussions in artificial intelligence applications to education and highlights the critical need for reliable, data-driven behavioral insights. We are the first to examine algorithmic recommendation adoption among adolescents and the factors driving aversion to algorithmic recommendations in educational decisions. Our findings suggest that this aversion stems from concerns about an algorithm's intent, ability, and alignment with students' preferences, highlighting broader challenges in AI explainability and trust. These concerns underscore the need for more human-aligned recommendation systems and the strategic integration of human oversight. Our experimental approach provides a generalizable framework for addressing a wide range of questions about leveraging AI across domains.

References

AlfaVita (2024). Η Τεχνητή Νοημοσύνη Μπαίνει Επίσημα στο Ελληνικό σχολείο: Ποιο Είναι το Πρώτο Εργαλείο που θα δουν οι Μαθητές (Artificial Intelligence Officially Enters the Greek School System: What Is the First Tool Students Will See?). Accessed: 2025-04-17.

Allman, M., I. Ashlagi, I. Lo, J. Love, K. Mentzer, L. Ruiz-Setz, and H. O'Connell (2022). Designing School Choice for Diversity in the San Francisco Unified School District. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pp. 290–291.

Arteaga, F., A. J. Kapor, C. A. Neilson, and S. D. Zimmerman (2022). Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice. *The Quarterly Journal of Economics 137*(3), 1791–1848.

Athey, S. and S. Wager (2019). Estimating Treatment Effects With Causal Forests: An Application. *Observational Studies 5*(2), 37–51.

Athey, S. and S. Wager (2021). Policy Learning With Observational Data. *Econometrica 89*(1), 133–161.

Baron, R. M. and D. A. Kenny (1986). The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology 51*(6), 1173–1182.

Bartmann, M. (2023). Reasoning with Recommender Systems?: Practical Reasoning, Digital Nudging, and Autonomy. In S. Genovesi, K. Kaesling, and S. Robbins (Eds.), *Recommender Systems: Legal and Ethical Issues*, pp. 129–145. Springer International Publishing.

Bastani, H., O. Bastani, A. Sungu, H. Ge, O. Kabakcı, and R. Mariman (2024). Generative AI Can Harm Learning. *Available at SSRN 4895486*.

Boatsman, J. R., C. Moeckel, and B. K. Pei (1997). The Effects of Decision Consequences on Auditors' Reliance on Decision Aids in Audit Planning. *Organizational Behavior and Human Decision Processes 71*(2), 211–247.

Bobba, M. and V. Frisancho (2022). Self-Perceptions About Academic Achievement: Evidence from Mexico City. *Journal of Econometrics 231*(1), 58–73.

Bogert, E., A. Schecter, and R. T. Watson (2021). Humans Rely More on Algorithms Than Social Influence as a Task Becomes More Difficult. *Scientific Reports 11*(1), 8028.

Bonaccio, S. and R. S. Dalal (2006). Advice Taking and Decision-Making: An Integrative Literature Review, and Implications for the Organizational Sciences. *Organizational Behavior and Human Decision Processes 101*(2), 127–151.

Burton, J. W., M. Stein, and T. B. Jensen (2020). A Systematic Review of Algorithm Aversion in Augmented Decision Making. *Journal of Behavioral Decision Making 33*(2), 220–239.

Cadario, R., C. Longoni, and C. K. Morewedge (2021). Understanding, Explaining, and Utilizing Medical Artificial Intelligence. *Nature Human Behaviour 5*(12), 1636–1642.

Castelo, N., M. W. Bos, and D. R. Lehmann (2019). Task-dependent Algorithm Aversion. *Journal of Marketing Research 56*(5), 809–825.

Choung, H., P. David, and A. Ross (2023). Trust in AI and its Role in the Acceptance of AI Technologies. *International Journal of Human–Computer Interaction 39*(9), 1727–1739.

Claybourn, C. (2023). Is AI Affecting College Admissions? *U.S. News*.

Connecticut Commission for Educational Technology (2024). 2024-2028 Strategic Plan for the Commission for Educational Technology. https://portal.ct.gov/-/media/das/ctedtech/publications/2024/202428_cet_strategic_plan.pdf.

Dargnies, M.-P., R. Hakimov, and D. Kübler (2024). Aversion to Hiring Algorithms: Transparency, Gender Profiling, and Self-confidence. *Management Science*.

DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2013). The Importance of Being Marginal: Gender Differences in Generosity. *American Economic Review 103*(3), 586–590.

Dietvorst, B. J., J. P. Simmons, and C. Massey (2015). Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. *Journal of Experimental Psychology: General 144*(1), 114–126.

Dietvorst, B. J., J. P. Simmons, and C. Massey (2018). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science 64*(3), 1155–1170.

Dusseault, B. and J. Lee (2023). AI is Already Disrupting Education, But Only 13 States Are Offering Guidance for Schools. Center on Reinventing Public Education (CRPE).

Eurostat (2024). Tertiary Education Statistics. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tertiary_education_statistics.

Faenza, Y., S. Gupta, and X. Zhang (2020). Reducing the Feeder Effect in Public School Admissions: A Bias-Aware Analysis for Targeted Interventions. *arXiv preprint arXiv:2004.10846*.

Fernández-Loría, C. and F. Provost (2022). Causal Classification: Treatment Effect Estimation Vs. Outcome Prediction. *Journal of Machine Learning Research 23*(59), 1–35.

Force, N. T. (2024). Report of the NEA Task Force on Artificial Intelligence in Education. Technical report, National Education Association.

Garg, N., H. Li, and F. Monachou (2020). Dropping Standardized Testing for Admissions Trades Off Information and Access. *arXiv preprint arXiv:2010.04396*.

Gedrimiene, E., I. Celik, A. Kaasila, K. Mäkitalo, and H. Muukkonen (2023, November). Artificial Intelligence (AI)-Enhanced Learning Analytics (LA) for Supporting Career Decisions: Advantages and Challenges from User Perspective. *Education and Information Technologies 29*(1), 297–322.

Gedrimiene, E., I. Celik, K. Mäkitalo, and H. Muukkonen (2023). Transparency and Trustworthiness in User Intentions to Follow Career Recommendations from a Learning Analytics Tool. *Journal of Learning Analytics 10*(1), 54–70.

Gibbons, J. L. and K. E. Poelker (2020). *Technology and the Global Adolescent.* Cambridge University Press.

Glikson, E. and A. W. Woolley (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals 14*(2), 627–660.

Goldfarb, A. and C. Tucker (2019). Digital Economics. *Journal of Economic Literature 57*(1), 3–43.

Goulas, S., S. Griselda, and R. Megalokonomou (2024). Comparative Advantage and Gender Gap in STEM. *Journal of Human Resources 59*(6), 1937–1980.

Goulas, S. and F. Monachou (2024). Optimal Cohort Partitions. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pp. 1097–1097.

Goyal, N., A. A. Okuno, D. Singhvi, and S. Singhvi (2024). Increasing Parents' Engagement on Ed-Tech Platforms: Evidence from the Field. *Available at SSRN 4918175*.

Highhouse, S. (2008). Stubborn Reliance on Intuition and Subjectivity in Employee Selection. *Industrial and Organizational Psychology 1*(3), 333–342.

Kang, H. and C. Lou (2022). Understanding Human–AI Interactions on TikTok and Their Implications for Algorithmic Literacy. *Journal of Computer-Mediated Communication 27*(5), zmac014.

Kathimerini (2023). Μάθημα τεχνητής νοημοσύνης, αλλαγές σε Αρχαία, Λογοτεχνία (artificial intelligence course, changes in ancient greek and literature). Accessed: 2025-04-14.

Kathimerini (2024). Η Τεχνητή Νοημοσύνη στο Γυμνάσιο (artificial intelligence in middle school). Accessed: 2025-04-14.

Keppler, S., W. P. Sinchaisri, and C. Snyder (2024). Backwards Planning with Generative AI: Case Study Evidence from US K12 Teachers. *Available at SSRN*.

Keppler, S. M., J. Li, and D. Wu (2022). Crowdfunding the Front Lines: An Empirical Study of Teacher-driven School Improvement. *Management Science 68*(12), 8809–8828.

Kizilcec, R. F. (2016). How Much Information?: Effects of Transparency on Trust in an Algorithmic Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose, California, USA, pp. 2390–2395. ACM.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2017). Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*.

Komiak, S. Y. X. and I. Benbasat (2006). The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. *MIS Quarterly 30*(4), 941–960.

Larroucau, T., I. Rios, A. Fabre, and C. Neilson (2024). College Application Mistakes and the Design of Information Policies at Scale. Mimeo.

Lee, M. K. (2018). Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management. *Big Data and Society 5*(1).

Lin, W., S.-H. Kim, and J. Tong (2021). Does Algorithm Aversion Exist in the Field? An Empirical Analysis of Algorithm Use Determinants in Diabetes Self-management. *SSRN Electron. J.*.

Logg, J. and R. Schlund (2024). A Simple Explanation Reconciles “Algorithm Aversion” and “Algorithm Appreciation”: Hypotheticals vs. Real Judgments. *Real Judgments (January 8, 2024)*.

Logg, J. M., J. A. Minson, and D. A. Moore (2019). Algorithm Appreciation: People Prefer Algorithmic to Human Judgment. *Organizational Behavior and Human Decision Processes 151*, 90–103.

Longoni, C., A. Bonezzi, and C. K. Morewedge (2019). Resistance to Medical Artificial Intelligence. *Journal of Consumer Research 46*(4), 629–650.

Longoni, C. and L. Cian (2020). Artificial Intelligence in Utilitarian vs. Hedonic Contexts: The ’Word-of-Machine’ Effect. *Journal of Marketing 86*(1), 91–108.

Moses-Payne, M. E., J. Habicht, A. Bowler, N. Steinbeis, and T. U. Hauser (2021). I Know Better! Emerging Metacognition Allows Adolescents to Ignore False Advice. *Developmental Science 24*(5), e13101.

Neilson, C. (2024). *World Development Report 2024: The Middle-Income Trap*. World Bank.

Newman, D. T., N. J. Fast, and D. J. Harmon (2020). When Eliminating Bias Isn’t Fair: Algorithmic Reductionism and Procedural Justice in Human Resource Decisions. *Organizational Behavior and Human Decision Processes 160*, 149–167.

Organisation for Economic Co-operation and Development (2018). *Education For A Bright Future In Greece*. OECD Publishing, Paris.

Prahl, A. and L. Van Swol (2017). Understanding Algorithm Aversion: When is Advice From Automation Discounted? *Journal of Forecasting 36*(6), 691–702.

Saavedraez, J. and E. Molina (2024). From Fear to Opportunity: Making AI Work for Education. https://blogs.worldbank.org/en/education/from-fear-to-opportunity--making-ai-work-for-education.

Shoval, R., N. Karsh, and B. Eitam (2022). Choosing to Choose or Not. *Judgment and Decision Making 17*(4), 768–796.

Sidoti, O., E. Park, and J. Gottfried (2025). Share of Teens Using ChatGPT for Schoolwork Doubled from 2023 to 2024. *Pew Research Center*.

Sileo, D. (2024). tasksource: A Large Collection of NLP Tasks with a Structured Dataset Preprocessing Framework. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 15655–15684.

Smilowitz, K. and S. Keppler (2020). On the Use of Operations Research and Management in Public Education Systems. *Pushing the Boundaries: Frontiers in Impactful OR/OM Research*, 84–105.

Spector, C. (2024). Technology in Education: Transforming the Future of Learning. https://news.stanford.edu/stories/2024/02/technology-in-education.

Starke, A., M. Willemsen, and C. Snijders (2021). Promoting Energy-Efficient Behavior by Depicting Social Norms in a Recommender Interface. *ACM Transactions on Interactive Intelligent Systems (TiiS) 11*(3-4), 1–32.

Sunstein, C. R. and J. Gaffe (2024). An Anatomy of Algorithm Aversion. *Available at SSRN*.

Tsitsika, A. K., E. C. Tzavela, M. Janikian, K. Ólafsson, A. Iordache, T. M. Schoenmakers, C. Tzavara, and C. Richardson (2014). Online Social Networking in Adolescence: Patterns of Use in Six European Countries and Links with Psychosocial Functioning. *Journal of Adolescent Health 55*(1), 141–147.

Varsik, S. and L. Vosberg (2024). The Impact of Digital Technologies on Education: Trends and Challenges. *OECD Publishing*.

Vodrahalli, K., R. Daneshjou, T. Gerstenberg, and J. Zou (2022). Do Humans Trust Advice More if it Comes from AI?: An Analysis of Human-AI Interactions. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, Oxford, United Kingdom, pp. 763–777. ACM.

Xu, L., Z. A. Pardos, and A. Pai (2023). Convincing the Expert: Reducing Algorithm Aversion in Administrative Higher Education Decision-Making. In *Proceedings of the Tenth ACM Conference on Learning@ Scale*, pp. 215–225.

Yeomans, M., A. Shah, S. Mullainathan, and J. Kleinberg (2019). Making Sense of Recommendations. *Journal of Behavioral Decision Making 32*(4), 403–414.

Zerilli, J., U. Bhatt, and A. Weller (2022). How Transparency Modulates Trust in Artificial Intelligence. *Patterns 3*(4), 100455.

Zhu, T., K. Zhang, and W. Y. Wang (2024, November). Embracing AI in Education: Understanding the Surge in Large Language Model Use by Secondary Students.

Supplementary Appendix

Students Choose Human Counselors Over Algorithms in College Applications, but Not Always

S1 Questionnaires

S1.1 Human Counselor

Q1. What is your grade level and class number? ___________

Q2. What is your gender identity? (Mark only one answer.)

☐ Female ☐ Male ☐ Other ☐ Prefer to not answer

Q3. What is your track choice, or what track choice are you planning to make? (Mark only one answer.)

☐ Humanities ☐ Life Sciences ☐ Exact Sciences ☐ Economics and IT

Q4. Do you intend to participate in the national exams, Panelladikes? (Mark only one answer.)

☐ Yes ☐ No ☐ I have not decided yet

Q5. What was your final GPA in the previous academic year? ______/20

Q6. What were your scores in the previous final exams (June 2024)?

- General-track Mathematics/ Algebra: ___________
- Greek Language: ___________

For the questions that follow, imagine the following hypothetical situation.

After the Panelladikes exams, candidates submit an ordered preference list of their preferred college programs to the Ministry of Education. Every year, the Ministry calculates the score (admission threshold) that a student needs to have to enter each school. If the student's grades are good enough to get into more than one school on their list, they are admitted to the school that is the highest on their list.

The Ministry wants to help high school students choose the right schools for them. Therefore, there will be **career counselors** in each school who will provide recommendations to the students. These counselors have worked with many students and professionals in Greece.

Three high school students, Eleni, Maria, and Anna, asked for help from these counselors to fill out their preference lists after the Panelladikes exams.

Eleni's story:

Eleni shared the following information with the counselor:

- She has scored **17,800** in the Panelladikes exams.
- She is ambitious and hard-working.
- She loves technology and has learned to code on her own.
- She lives in Heraklion in Crete with her family, and has no family in Athens.
- She wants to list the following schools in this order:

1. Informatics at Athens University of Economics and Business (Last year's Threshold: 17,220)
2. Computer Science at University of Crete - Heraklion (Last year's Threshold: 15,813)

Using students' home location as criterion, the counselor suggested the following order of programs (#1 being highest priority):

1. Computer Science at the University of Crete - Heraklion (Last year's Threshold: 15,813)
2. Informatics at the Athens University of Economics and Business (Last year's Threshold: 17,220)

Please mark your opinion (X on the opinion line) for the following statements:

Q7. I trust the counselor has Eleni's best interest in mind.

-5		0		+5
Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree

Q8. I believe the counselor knows more about college choices than Eleni.

-5		0		+5
Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree

Q9. I understand why the counselor made this recommendation to Eleni.

-5		0		+5
Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree

Q10. I believe the counselor's recommendation is compatible with Eleni's preferences.

-5		0		+5
Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree

Q11. If you were Eleni, what would be your final decision? Please number the options as #1 and #2 (with #1 as your top priority).

☐ Informatics at the Athens University of Economics and Business (Last year's Threshold: 17,220)

☐ Computer Science at the University of Crete - Heraklion (Last year's Threshold: 15,813)

Maria's story:

Maria shared the following information with the counselor:

- She has scored **18,500** in the Panelladikes exams.
- She is ambitious and hard-working.
- Her favorite subject is History and she regularly posts Archaeology explainers on her Youtube channel with millions of views.
- She resides in Athens with her family.
- She wants to list the following schools in this order:

1. Archaeology at the University of Athens (Last year's Threshold: 11,958)
2. Law School at the University of Athens (Last year's Threshold: 18,025)

Using the programs' popularity reflected in last year's threshold as criterion, the counselor suggested the following order of programs (#1 being highest priority):

1. Law School at the University of Athens (Last year's Threshold: 18,025)
2. Archaeology at the University of Athens (Last year's Threshold: 11,958)

Please mark your opinion (X on the opinion line) for the following statements:

Q12. I trust the counselor has Maria's best interest in mind.

-5 O +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q13. I believe the counselor knows more about college choices than Maria.

-5 O +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q14. I understand why the counselor made this recommendation to Maria.

-5 O +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q15. I believe the counselor's recommendation is compatible with Maria's preferences.

-5 O +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q16. If you were Maria, what would be your final decision? Please number the options as #1 and #2 (with #1 as your top preference).

☐ Archaeology at the University of Athens (Last year's Threshold: 11,958)

☐ Law School at the University of Athens (Last year's Threshold: 18,025)

Anna's story:

Anna shared the following information with the counselor:

- She has scored **17,700** in the Panelladikes exams.
- She's ambitious and hard-working.
- Anna has always dreamed of becoming a doctor and is interested in biological sciences.
- She is not sure about her future career but wants to study something she will enjoy.
- She lives in Crete with her family but is willing to move for her studies.
- She wants to list the following schools in this order:

1. Medicine at University of Athens (Last year's Threshold: 18,775)
2. Medicine at University of Thessaly (Last year's Threshold: 18,125)

She is also considering adding one of the following options as her third choice:

- Medicine at the University of Crete (Last year's Threshold: 18,050)
- Pharmacy at the University of Patras (Last year's Threshold: 17,760)
- Molecular Biology and Genetics at the University of Thrace (Last year's Threshold: 16,570)

Using the admission chances as criterion, the counselor suggested the following as the third choice:

3. Molecular Biology and Genetics at the University of Thrace (Last year's Threshold: 16,570)

Please mark your opinion (X on the opinion line) for the following statements:

Q17. I trust the counselor has Anna's best interest in mind.

-5 | O | +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q18. I trust the counselor knows more about college choices than Anna.

-5 | O | +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q19. I understand why the counselor made this recommendation to Anna.

-5 | O | +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q20. I believe the counselor's recommendation is compatible with Anna's preferences.

-5 | O | +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q21. As Anna, what would be your final decision? (Mark only one answer.)

- ☐ I would choose Medicine at the University of Crete as the third choice (Last year's Threshold: 18,050).
- ☐ I would choose Pharmacy at the University of Patras as the third choice (Last year's Threshold: 17,670).
- ☐ I would choose Molecular Biology and Genetics at the University of Thrace as the third choice (Last year's Threshold: 16,570).
- ☐ I would make no changes to Anna's original preference list.

The following questions are about your own college program choices.

Q22. I will apply to the most competitive programs I have a chance of being admitted to regardless of my other interests.

-5 | O | +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q23. I would like to receive personalized recommendations on my college program choices from a counselor like the one in the stories.

-5 | O | +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q24. How could such a counselor be more helpful to you?

Q25. What are your interests? For e.g, you might like painting, sports, or mathematics.

Q26. Please rank the following disciplines based on your interests:

- ☐ Engineering and Informatics
- ☐ Health Sciences
- ☐ Law and Humanities
- ☐ Economics
- ☐ Other: ________

S1.2 Algorithmic Recommender

Q1. What is your grade level and class number? __________
Q2. What is your gender identity? (Mark only one answer.)

☐ Female ☐ Male ☐ Other ☐ Prefer to not answer

Q3. What is your track choice, or what track choice are you planning to make? (Mark only one answer.)

☐ Humanities ☐ Life Sciences ☐ Exact Sciences ☐ Economics and IT

Q4. Do you intend to participate in the national exams, Panelladikes? (Mark only one answer.)

☐ Yes ☐ No ☐ I have not decided yet

Q5. What was your final GPA in the previous academic year? _____/20
Q6. What were your scores in the previous final exams (June 2024)?

- General-track Mathematics/ Algebra: __________
- Greek Language: __________

For the questions that follow, imagine the following hypothetical situation.

After the Panelladikes exams, candidates submit an ordered preference list of their preferred college programs to the Ministry of Education. Every year, the Ministry calculates the score (admission threshold) that a student needs to have to enter each school. If the student's grades are good enough to get into more than one school on their list, they are admitted to the school that is the highest on their list.

The Ministry wants to help high school students choose the right schools for them. Therefore, there will be **an AI algorithmic system ("algorithm")** available in each school's computer lab to provide recommendations to the students. This algorithm uses data from students across all of Greece.

Three high school students, Eleni, Maria, and Anna, asked for help from these counselors to fill out their preference lists after the Panelladikes exams.

Eleni's story:
Eleni shared the following information with the counselor:

- She has scored **17,800** in the Panelladikes exams.
- She is ambitious and hard-working.
- She loves technology and has learned to code on her own.
- She lives in Heraklion in Crete with her family, and has no family in Athens.
- She wants to list the following schools in this order:

1. Informatics at Athens University of Economics and Business (Last year's Threshold: 17,220)
2. Computer Science at University of Crete - Heraklion (Last year's Threshold: 15,813)

Using students' home location as criterion, the algorithm suggested the following order of programs (#1 being highest priority):

1. Computer Science at the University of Crete - Heraklion (Last year's Threshold: 15,813)
2. Informatics at the Athens University of Economics and Business (Last year's Threshold: 17,220)

Please mark your opinion (X on the opinion line) for the following statements:
Q7. I trust the algorithm has Eleni's best interest in mind.

-5		O		+5
Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree

Q8. I believe the algorithm knows more about college choices than Eleni.

-5		O		+5
Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree

Q9. I understand why the algorithm made this recommendation to Eleni.

-5		O		+5
Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree

Q10. I believe the algorithm's recommendation is compatible with Eleni's preferences.

-5		O		+5
Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree

Q11. If you were Eleni, what would be your final decision? Please number the options as #1 and #2 (with #1 as your top priority).

☐ Informatics at the Athens University of Economics and Business (Last year's Threshold: 17,220)

☐ Computer Science at the University of Crete - Heraklion (Last year's Threshold: 15,813)

Maria's story:

Maria shared the following information with the counselor:

- She has scored **18,500** in the Panelladikes exams.
- She is ambitious and hard-working.
- Her favorite subject is History and she regularly posts Archaeology explainers on her Youtube channel with millions of views.
- She resides in Athens with her family.
- She wants to list the following schools in this order:

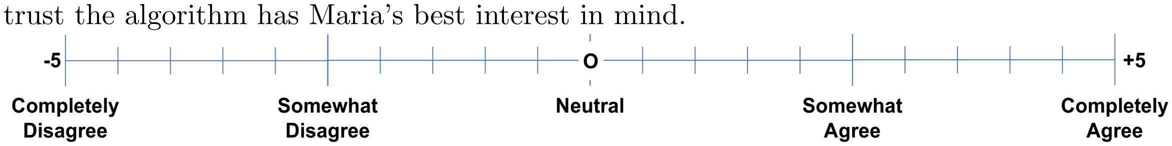
1. Archaeology at the University of Athens (Last year's Threshold: 11,958)
2. Law School at the University of Athens (Last year's Threshold: 18,025)

Using the programs' popularity reflected in last year's threshold as criterion, the algorithm suggested the following order of programs (#1 being highest priority):

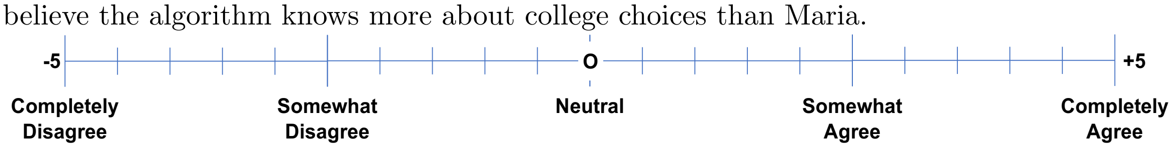
1. Law School at the University of Athens (Last year's Threshold: 18,025)
2. Archaeology at the University of Athens (Last year's Threshold: 11,958)

Please mark your opinion (X on the opinion line) for the following statements:

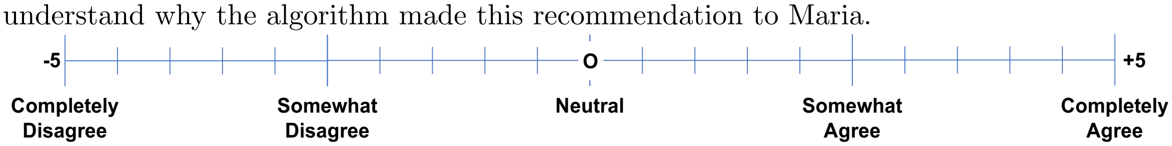
Q12. I trust the algorithm has Maria's best interest in mind.



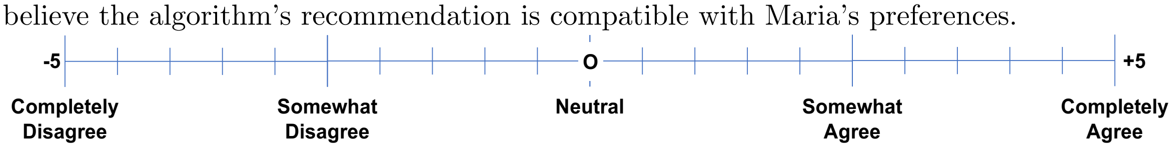
Q13. I believe the algorithm knows more about college choices than Maria.



Q14. I understand why the algorithm made this recommendation to Maria.



Q15. I believe the algorithm's recommendation is compatible with Maria's preferences.



Q16. If you were Maria, what would be your final decision? Please number the options as #1 and #2 (with #1 as your top preference).

☐ Archaeology at the University of Athens (Last year's Threshold: 11,958)

☐ Law School at the University of Athens (Last year's Threshold: 18,025)

Anna's story:

Anna shared the following information with the counselor:

- She has scored **17,700** in the Panelladikes exams.
- She's ambitious and hard-working.
- Anna has always dreamed of becoming a doctor and is interested in biological sciences.
- She is not sure about her future career but wants to study something she will enjoy.
- She lives in Crete with her family but is willing to move for her studies.
- She wants to list the following schools in this order:

1. Medicine at University of Athens (Last year's Threshold: 18,775)
2. Medicine at University of Thessaly (Last year's Threshold: 18,125)

She is also considering adding one of the following options as her third choice:

- Medicine at the University of Crete (Last year's Threshold: 18,050)
- Pharmacy at the University of Patras (Last year's Threshold: 17,760)
- Molecular Biology and Genetics at the University of Thrace (Last year's Threshold: 16,570)

Using the admission chances as criterion, the algorithm suggested the following as the third choice:

3. Molecular Biology and Genetics at the University of Thrace (Last year's Threshold: 16,570)

Please mark your opinion (X on the opinion line) for the following statements:

Q17. I trust the algorithm has Anna's best interest in mind.

-5 O +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q18. I trust the algorithm knows more about college choices than Anna.

-5 O +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q19. I understand why the algorithm made this recommendation to Anna.

-5 O +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q20. I believe the algorithm's recommendation is compatible with Anna's preferences.

-5 O +5

Completely Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Completely Agree

Q21. As Anna, what would be your final decision? (Mark only one answer.)

☐ I would choose Medicine at the University of Crete as the third choice (Last year's Threshold: 18,050).

☐ I would choose Pharmacy at the University of Patras as the third choice (Last year's Threshold: 17,670).

☐ I would choose Molecular Biology and Genetics at the University of Thrace as the third choice (Last year's Threshold: 16,570).

☐ I would make no changes to Anna's original preference list.

The following questions are about your own college program choices.

Q22. I will apply to the most competitive programs I have a chance of being admitted to regardless of my other interests.

-5		0		+5
Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree

Q23. I would like to receive personalized recommendations on my college program choices from an algorithmic counselor like the ones in the stories.

-5		0		+5
Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree

Q24. How could such a algorithmic counselor be more helpful to you?

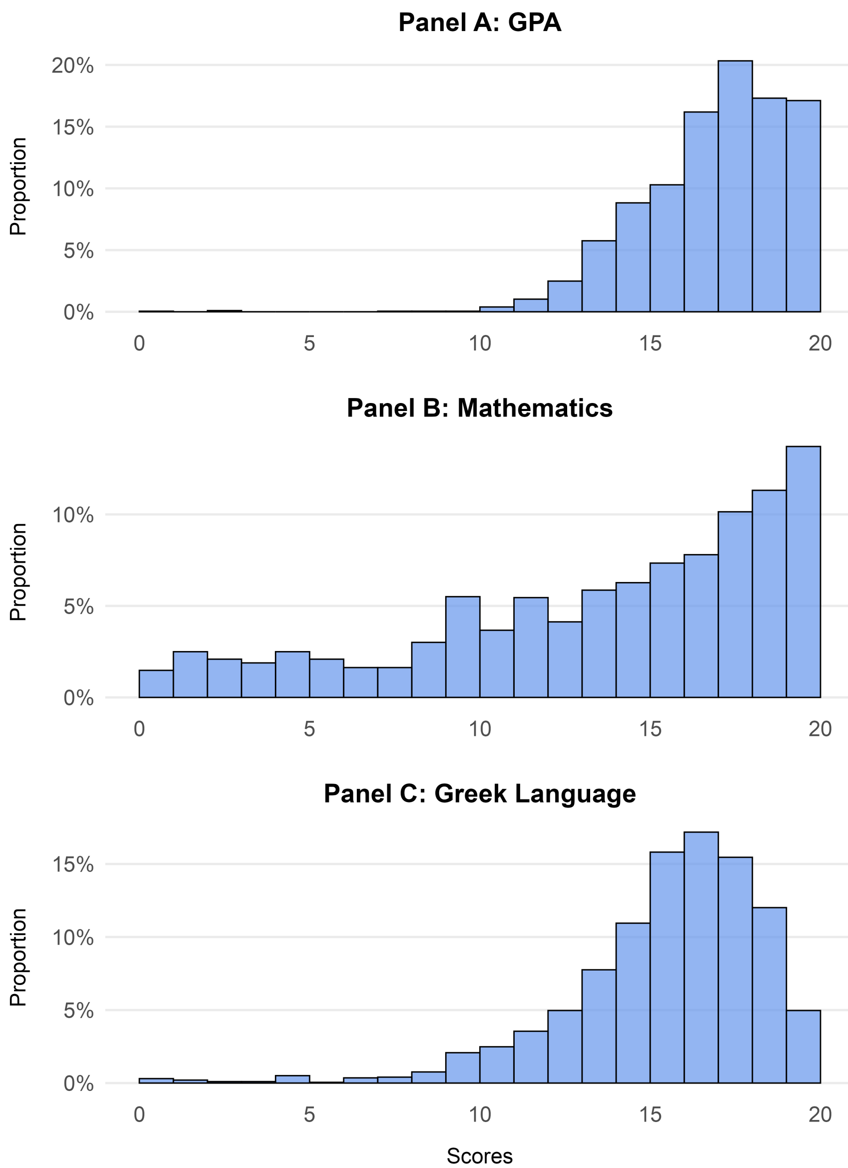
Q25. What are your interests? For e.g, you might like painting, sports, or mathematics.

Q26. Please rank the following disciplines based on your interests:

☐ Engineering and Informatics ☐ Health Sciences ☐ Law and Humanities ☐ Economics ☐ Other: ________

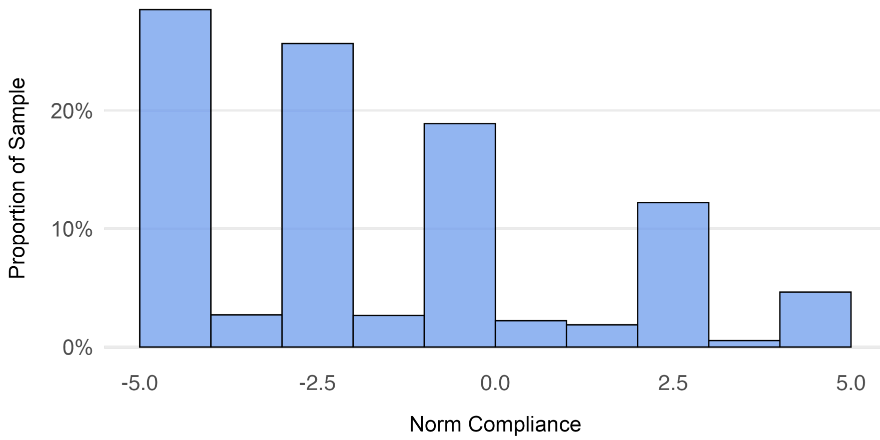
S2 Descriptive Statistics

Figure S1: Distribution of Prior-year Student Performance



Notes: Panel A displays the histogram of students' reported prior-year Grade Point Average (GPA). Panel B shows the histogram of reported mathematics scores from the prior-year final exams. Panel C illustrates the histogram of reported Greek language scores from the prior-year final exams.

Figure S2: Distribution of Norm Compliance Score



Notes: This figure shows the histogram of the norm compliance score. *Norm compliance* is measured on a scale from -5 to 5, based on students' agreement with the statement: *"I will apply to the most competitive programs I have a chance of being admitted to regardless of my other interests."*

Table S1: Representativeness of Sampled High Schools

	Sample (14 Schools) Mean	Population (1,223 Schools) Mean	Difference (s.e.)
Student Characteristics			
Female (1=Yes)	0.56	0.58	-0.015
			(0.013)
Age (Years)	17.96	17.93	0.027
			(0.018)
Born in 1st Quarter of Year (1=Yes)	0.13	0.13	-0.002
			(0.007)
Tertiary Education Enrollment (1=Yes)	0.75	0.78	-0.029
			(0.024)
Admitted to Higher Educational Institutions (1=Yes)	0.53	0.57	-0.036
			(0.029)
Apply to STEM Degree Programs (1=Yes)	0.62	0.63	-0.016
			(0.013)
University Admission Score (/20,000)	13,358.38	13,587.04	-228.664
			(291.163)
Track Choice (1=Yes):			
Humanities Track	0.43	0.40	0.031
			(0.019)
Science Track	0.12	0.16	-0.036
			(0.020)
Information Technology	0.45	0.45	0.005
			(0.020)
School Characteristics			
Postcode Income (in 2009 Euros, Annual)	19,512.31	18,359.62	1,152.687
			(778.481)
Rural (1=Yes)	0.24	0.29	-0.044
			(0.126)

Notes: The table reports the differences in student and school characteristics between the schools in the sample used in this analysis (column 1) and all remaining schools in Greece (column 2), along with the standard errors for the significance of these differences (column 3). Data come from the Ministry of Education for all students across the nation in the graduating cohorts between 2003 and 2011. The population includes all traditional and experimental public schools; evening, private, and special schools are excluded. Track structure has changed as shown in Table 1. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

S3 Main Estimates and Heterogeneous Treatment Effects

Table S2: Estimated Recommendation Adoption Rates, Logistic Regression

	Means		Without Controls		With Controls		
Scenario	Human	Algorithm	$\hat{\beta}$	SE	$\hat{\beta}$	SE	N
Heart	0.315	0.318	-0.003	0.024	-0.008	0.023	2,048
Geography	0.365	0.343	-0.030*	0.017	-0.030*	0.016	2,051
Pragmatism	0.471	0.419	-0.050**	0.021	-0.047**	0.023	2,039

Notes: Parameter $\hat{\beta}$ is the estimated marginal effect from a logistic regression model. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S3: Estimated Choice Adoption In Pragmatism Scenario

	Means		With Controls	
Choice	Human	Algorithm	$\hat{\beta}$	SE
Medicine	0.197	0.223	0.023	0.018
Pharmacy	0.236	0.243	0.008	0.019
Molecular Biology *(Recommendation)*	0.471	0.419	-0.047**	0.023
None of the Above	0.095	0.114	0.016	0.014

Notes: The table presents the estimated choice adoption rates in the Pragmatism scenario. Parameter $\hat{\beta}$ is the estimated marginal effect from a linear probability model. In the pragmatism scenario, respondents had four choice options and the recommendation source (Human Counselor or Algorithmic Recommender) suggested the third option (Molecular Biology). Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S4: Heterogeneous Recommendation Adoption Rate Gap by Gender

	Scenario					
	Heart		Geography		Pragmatism	
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	-0.030	-0.039*	-0.030	-0.027	-0.065	-0.065
	(0.022)	(0.022)	(0.030)	(0.028)	(0.039)	(0.040)
Algorithm × Female	0.034	0.047	-0.003	-0.003	0.018	0.016
	(0.036)	(0.035)	(0.053)	(0.049)	(0.034)	(0.035)
Observations	2,048	2,048	2,057	2,057	2,039	2,039
Controls		✓		✓		✓

Notes: Algorithm represents the estimated marginal effect of assignment to the algorithmic recommender condition in a linear probability model, with the human counselor condition as the reference group. *Female* is a binary indicator that equals one if the respondent reports being female and zero if male. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S5: Heterogeneous Recommendation Adoption Rate Gap by GPA Tertile

	Scenario					
	Heart		Geography		Pragmatism	
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	0.015	0.008	-0.027	-0.022	-0.082**	-0.080**
	(0.051)	(0.050)	(0.019)	(0.020)	(0.034)	(0.033)
Algorithm × Middle Tertile [1]	-0.009	-0.012	0.025	0.018	0.018	0.021
	(0.042)	(0.043)	(0.043)	(0.045)	(0.046)	(0.048)
Algorithm × Top Tertile [2]	-0.054	-0.037	-0.040	-0.048*	0.078*	0.074*
	(0.058)	(0.056)	(0.024)	(0.024)	(0.039)	(0.036)
Observations	2,048	2,048	2,057	2,057	2,039	2,039
P-value for H0: [1] = [2]	0.328	0.600	0.144	0.168	0.273	0.346
Controls		✓		✓		✓

Notes: Algorithm represents the estimated marginal effect of assignment to the algorithmic recommender condition in a linear probability model, with the human counselor condition as the reference group. *Middle Tertile* is a binary indicator equal to one if the respondent's prior-year GPA falls in the middle tertile of the sample distribution and zero otherwise. *Top Tertile* is defined similarly for the top tertile of the GPA distribution. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S6: Heterogeneous Recommendation Adoption Rate Gap by Norm Compliance Tertile

	Scenario					
	Heart		Geography		Pragmatism	
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	0.011	0.005	-0.031	-0.033	-0.014	-0.017
	(0.038)	(0.037)	(0.043)	(0.041)	(0.033)	(0.037)
Algorithm × Middle Tertile [1]	-0.006	-0.003	0.038	0.040	-0.051	-0.043
	(0.048)	(0.046)	(0.049)	(0.046)	(0.039)	(0.041)
Algorithm × Top Tertile [2]	-0.064	-0.058	-0.100	-0.091	-0.062	-0.057
	(0.079)	(0.077)	(0.076)	(0.072)	(0.048)	(0.050)
Observations	2,048	2,048	2,057	2,057	2,039	2,039
P-value for H0: [1] = [2]	0.395	0.422	0.023	0.027	0.820	0.783
Controls		✓		✓		✓

Notes: Algorithm represents the estimated marginal effect of assignment to the algorithmic recommender condition in a linear probability model, with the human counselor condition as the reference group. *Middle Tertile* is a binary indicator equal to one if the respondent's norm compliance falls in the middle tertile of the sample distribution and zero otherwise. *Top Tertile* is defined similarly for the top tertile of the norm compliance distribution. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S7: Heterogeneous Recommendation Adoption Rate Gap by Urbanicity

	Scenario					
	Heart		Geography		Pragmatism	
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	-0.023	-0.024	-0.030	-0.030	-0.035	-0.033
	(0.024)	(0.022)	(0.018)	(0.017)	(0.023)	(0.023)
Algorithm × Rural	0.076	0.069	0.001	0.002	-0.058	-0.060
	(0.061)	(0.057)	(0.040)	(0.048)	(0.046)	(0.047)
Observations	2,048	2,048	2,057	2,057	2,039	2,039
Controls		✓		✓		✓

Notes: Algorithm represents the estimated marginal effect of assignment to the algorithmic recommender condition in a linear probability model, with the human counselor condition as the reference group. *Rural* is a binary indicator that equals one for respondents in rural schools and zero otherwise. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S8: Heterogeneous Recommendation Adoption Rate Gap by College Aspiration

	Scenario					
	Heart		Geography		Pragmatism	
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	0.003	0.007	-0.127***	-0.124***	-0.034	-0.037
	(0.060)	(0.059)	(0.030)	(0.033)	(0.071)	(0.069)
Algorithm × Exam Participation: Yes	-0.007	-0.016	0.117***	0.112**	-0.020	-0.013
	(0.055)	(0.056)	(0.038)	(0.040)	(0.072)	(0.066)
Observations	2,048	2,048	2,057	2,057	2,039	2,039
Controls		✓		✓		✓

Notes: Algorithm represents the estimated marginal effect of assignment to the algorithmic recommender condition in a linear probability model, with the human counselor condition as the reference group. *Exam Participation: Yes* is a binary indicator that equals one if the respondent reports an intention to participate in the national exams for university admission and zero otherwise. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S9: Heterogeneity in Recommendation Adoption by Intention to Use Recommender

	Means		Without Controls		With Controls		
Scenario	Human	Algorithm	$\hat{\beta}$	SE	$\hat{\beta}$	SE	N
By Intention to Use Recommender							
Heart							
Yes	0.305	0.332	0.021	0.024	0.015	0.023	1,612
No	0.391	0.282	-0.114**	0.048	-0.117**	0.048	394
Geography							
Yes	0.367	0.353	-0.022	0.024	-0.022	0.024	1,613
No	0.331	0.282	-0.056	0.040	-0.051	0.039	398
Pragmatism							
Yes	0.490	0.442	-0.041**	0.020	-0.038*	0.021	1,619
No	0.376	0.350	-0.034	0.060	-0.041	0.061	397

Notes: Parameter $\hat{\beta}$ is the estimated marginal effect from a linear probability model. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S10: Heterogeneity in Recommendation Adoption by Grade

	Means		Without Controls		With Controls		
Scenario	Human	Algorithm	$\hat{\beta}$	SE	$\hat{\beta}$	SE	N
By Grade							
Heart							
10	0.378	0.333	-0.051	0.046	-0.053	0.048	718
11	0.295	0.346	0.042	0.029	0.044*	0.026	704
12	0.256	0.270	0.014	0.046	0.009	0.044	602
Geography							
10	0.385	0.329	-0.074***	0.023	-0.082***	0.023	724
11	0.354	0.345	-0.013	0.026	-0.007	0.025	706
12	0.353	0.353	-0.003	0.040	-0.002	0.043	604
Pragmatism							
10	0.455	0.408	-0.039	0.041	-0.029	0.043	713
11	0.488	0.424	-0.060**	0.027	-0.061**	0.031	700
12	0.471	0.421	-0.054	0.049	-0.058	0.049	605

Notes: Parameter $\hat{\beta}$ is the estimated marginal effect from a linear probability model. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S11: Heterogeneity in Recommendation Adoption by Track

	Means		Without Controls		With Controls		
Scenario	Human	Algorithm	$\hat{\beta}$	SE	$\hat{\beta}$	SE	N
By Track							
Heart							
Humanities	0.316	0.345	0.013	0.045	0.005	0.045	564
Life Sciences	0.344	0.259	-0.082*	0.047	-0.072	0.047	417
Exact Sciences	0.317	0.305	-0.018	0.050	-0.020	0.043	478
Economics and IT	0.294	0.349	0.053	0.038	0.045	0.038	547
Geography							
Humanities	0.336	0.332	-0.017	0.050	-0.016	0.053	566
Life Sciences	0.338	0.294	-0.045	0.047	-0.043	0.046	417
Exact Sciences	0.378	0.374	-0.012	0.047	-0.015	0.044	481
Economics and IT	0.397	0.368	-0.031	0.031	-0.031	0.034	552
Pragmatism							
Humanities	0.478	0.426	-0.036	0.049	-0.033	0.050	558
Life Sciences	0.507	0.456	-0.057	0.062	-0.056	0.061	415
Exact Sciences	0.478	0.416	-0.061	0.055	-0.059	0.057	474
Economics and IT	0.433	0.394	-0.043	0.044	-0.044	0.043	550

Notes: Parameter $\hat{\beta}$ is the estimated marginal effect from a linear probability model. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S12: Heterogeneity in Recommendation Adoption by Prior-year Performance (Mathematics + Greek Language)

	Means		Without Controls		With Controls		
Scenario	Human	Algorithm	$\hat{\beta}$	SE	$\hat{\beta}$	SE	N
By Tertile of Prior-year Performance (Mathematics + Greek Language)							
Heart							
Bottom	0.307	0.358	0.039	0.047	0.026	0.047	692
Middle	0.321	0.303	-0.020	0.029	-0.026	0.028	583
Top	0.307	0.286	-0.020	0.042	-0.013	0.042	644
Geography							
Bottom	0.391	0.359	-0.047*	0.028	-0.041	0.028	697
Middle	0.360	0.337	-0.022	0.045	-0.028	0.047	587
Top	0.327	0.321	-0.015	0.034	-0.020	0.034	645
Pragmatism							
Bottom	0.453	0.366	-0.079***	0.022	-0.080***	0.023	687
Middle	0.500	0.440	-0.063	0.054	-0.051	0.053	587
Top	0.495	0.472	-0.019	0.036	-0.016	0.036	639

Notes: Parameter $\hat{\beta}$ is the estimated marginal effect from a linear probability model. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S13: Heterogeneous Recommendation Adoption Rate Gap by Intention to Use Recommender

	Scenario					
	Heart		Geography		Pragmatism	
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	-0.114**	-0.117**	-0.056	-0.051	-0.034	-0.041
	(0.048)	(0.048)	(0.040)	(0.039)	(0.060)	(0.061)
Algorithm × Would Use Recommender	0.135***	0.133**	0.034	0.030	-0.008	0.003
	(0.044)	(0.046)	(0.055)	(0.055)	(0.057)	(0.056)
Observations	2,006	2,006	2,011	2,011	2,016	2,016
Controls		✓		✓		✓

Notes: Algorithm represents the estimated marginal effect of assignment to the algorithmic recommender condition in a linear probability model, with the human counselor condition as the reference group. *Would Use Recommender* is a binary indicator that equals one if the respondent reports an intention (score greater than or equal to zero) to use a recommender like the one depicted in the scenarios (algorithm or human) and zero otherwise. All specifications include the indicator *Would Use Recommender* as well as an indicator for missing values in the relevant question. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S14: Heterogeneous Recommendation Adoption Rate Gap by Grade

	Scenario					
	Heart		Geography		Pragmatism	
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	-0.051	-0.053	-0.074***	-0.082***	-0.039	-0.029
	(0.046)	(0.048)	(0.023)	(0.023)	(0.041)	(0.043)
Algorithm × Grade 11 [1]	0.093*	0.097*	0.062**	0.075**	-0.020	-0.032
	(0.050)	(0.050)	(0.028)	(0.029)	(0.058)	(0.063)
Algorithm × Grade 12 [2]	0.065	0.062	0.071	0.080	-0.014	-0.029
	(0.067)	(0.068)	(0.048)	(0.048)	(0.059)	(0.059)
Observations	2,048	2,048	2,057	2,057	2,039	2,039
P-value for H0: [1] = [2]	0.598	0.512	0.862	0.925	0.917	0.962
Controls		✓		✓		✓

Notes: Algorithm represents the estimated marginal effect of assignment to the algorithmic recommender condition in a linear probability model, with the human counselor condition as the reference group. *Grade 11* is a binary indicator equal to one if the respondent reported being in grade 11 and zero otherwise. *Grade 12* is defined similarly for respondents reporting being in grade 12. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S15: Heterogeneous Recommendation Adoption Rate Gap by Tertile of Prior-year Performance (Mathematics + Greek Language)

	Scenario					
	Heart		Geography		Pragmatism	
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	0.039	0.025	-0.047	-0.042	-0.079***	-0.080***
	(0.047)	(0.048)	(0.028)	(0.028)	(0.022)	(0.023)
Algorithm × Middle Tertile [1]	-0.060	-0.051	0.025	0.013	0.016	0.029
	(0.049)	(0.050)	(0.047)	(0.049)	(0.049)	(0.048)
Algorithm × Top Tertile [2]	-0.059	-0.038	0.033	0.021	0.060	0.064
	(0.069)	(0.068)	(0.049)	(0.047)	(0.040)	(0.037)
Observations	2,048	2,048	2,057	2,057	2,039	2,039
P-value for H0: [1] = [2]	0.991	0.824	0.914	0.911	0.524	0.592
Controls		✓		✓		✓

Notes: Algorithm represents the estimated marginal effect of assignment to the algorithmic recommender condition in a linear probability model, with the human counselor condition as the reference group. *Middle Tertile* is a binary indicator equal to one if the respondent's prior-year performance (Mathematics + Greek Language) falls in the middle tertile of the sample distribution and zero otherwise. *Top Tertile* is defined similarly for the top tertile of the performance distribution. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and prior-year performance tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S16: Heterogeneous Recommendation Adoption Rate Gap by Track

	Scenario					
	Heart		Geography		Pragmatism	
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	0.015	0.005	-0.048	-0.016	-0.026	-0.033
	(0.041)	(0.045)	(0.027)	(0.053)	(0.043)	(0.050)
Algorithm × Life Sciences [1]	-0.069	-0.078	-0.022	-0.026	0.002	-0.022
	(0.045)	(0.058)	(0.045)	(0.078)	(0.047)	(0.063)
Algorithm × Exact Sciences [2]	-0.035	-0.025	0.061*	0.001	-0.037	-0.026
	(0.051)	(0.057)	(0.030)	(0.063)	(0.057)	(0.078)
Algorithm × Economics and IT [3]	0.017	0.040	0.039	-0.014	-0.048	-0.010
	(0.039)	(0.050)	(0.038)	(0.079)	(0.053)	(0.075)
Observations	2,048	2,048	2,057	2,057	2,039	2,039
P-value for H0: [1] = [2]	0.551	0.523	0.161	0.727	0.420	0.966
P-value for H0: [1] = [3]	0.065	0.039	0.066	0.796	0.330	0.894
P-value for H0: [2] = [3]	0.377	0.330	0.613	0.776	0.762	0.834
Controls		✓		✓		✓

Notes: *Algorithm* represents the estimated marginal effect of assignment to the algorithmic recommender condition in a linear probability model, with the human counselor condition as the reference group. *Life Sciences* is a binary indicator equal to one if the respondent's track is Life Sciences and zero otherwise. *Exact Sciences* and *Economics and IT* are defined analogously for students in the Exact Sciences and Economics and IT tracks, respectively. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and prior-year performance tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

S4 Mediation Analysis

Table S17: Effect of Algorithmic Recommender on Perceived Mediators

	Outcome Variable			
	Intent	Ability	Comprehension	Alignment
	(1)	(2)	(3)	(4)
	Heart Scenario			
Algorithm	-0.353**	-1.250***	0.279*	0.194
	(0.128)	(0.128)	(0.140)	(0.139)
Observations	2,058	2,055	2,057	2,057
	Geography Scenario			
Algorithm	-0.578***	-1.876***	0.408***	-0.006
	(0.103)	(0.120)	(0.111)	(0.131)
Observations	2,061	2,059	2,056	2,051
	Pragmatism Scenario			
Algorithm	-0.668***	-1.199***	-0.042	-0.414***
	(0.090)	(0.108)	(0.067)	(0.097)
Observations	2,052	2,051	2,047	2,051

Notes: This table presents the estimated difference in mediator scores between students in the *algorithmic recommender* and *human counselor* conditions. *Algorithm* represents the estimated marginal effect of assignment to the algorithmic recommender condition in a linear probability model, with the human counselor condition as the reference group. All specifications include controls and school fixed effects. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S18: Mediated Recommendation Adoption Rates Gap

	Heart		Geography		Pragmatism	
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	-0.008	-0.003	-0.030*	-0.005	-0.047*	-0.015
	(0.023)	(0.016)	(0.015)	(0.021)	(0.023)	(0.024)
Mediators:						
Intent		0.021***		0.036***		0.017**
		(0.003)		(0.004)		(0.007)
Ability		0.010***		0.003		0.000
		(0.003)		(0.005)		(0.004)
Comprehension		0.014***		0.005		0.030***
		(0.003)		(0.005)		(0.004)
Alignment		0.056***		0.047***		0.047***
		(0.004)		(0.004)		(0.005)
Observations	2,048	2,048	2,057	2,057	2,039	2,039
Controls	✓	✓	✓	✓	✓	✓

Notes: Participants rated each mediating channel's associated statement on a scale ranging from -5 to 5. Estimates come from a linear probability model. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S19: Mediated Recommendation Adoption Rates Gap, Heart Scenario

	Outcome: Recommendation Adoption					
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	-0.008	-0.007	-0.014	0.005	0.023	-0.003
	(0.023)	(0.017)	(0.016)	(0.017)	(0.020)	(0.016)
Mediators:						
Intent			0.024***	0.026***	0.039***	0.021***
			(0.003)	(0.003)	(0.003)	(0.003)
Ability		0.015***		0.012***	0.019***	0.010***
		(0.002)		(0.003)	(0.002)	(0.003)
Comprehension		0.020***	0.015***		0.019***	0.014***
		(0.003)	(0.002)		(0.003)	(0.003)
Alignment		0.061***	0.057***	0.057***		0.056***
		(0.004)	(0.004)	(0.004)		(0.004)
Observations	2,048	2,048	2,048	2,048	2,048	2,048
Student Controls	✓	✓	✓	✓	✓	✓

Notes: This table presents the mediating effects of different mediator combinations on the impact of the algorithmic recommender on recommendation adoption in the *Heart* scenario. Each model iteratively excludes one mediator to assess the sensitivity of the remaining mediator coefficients. Participants rated each mediating channel's associated statement on a scale ranging from -5 to 5. Estimates come from a linear probability model. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S20: Mediated Recommendation Adoption Rates Gap, Geography Scenario

	Outcome: Recommendation Adoption					
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	-0.030*	-0.016	-0.011	-0.000	0.012	-0.005
	(0.015)	(0.020)	(0.017)	(0.023)	(0.022)	(0.021)
Mediators:						
Intent			0.037***	0.038***	0.055***	0.036***
			(0.004)	(0.004)	(0.003)	(0.004)
Ability		0.010*		0.004	0.008	0.003
		(0.005)		(0.005)	(0.005)	(0.005)
Comprehension		0.014***	0.005		0.011**	0.005
		(0.004)	(0.005)		(0.005)	(0.005)
Alignment		0.055***	0.047***	0.047***		0.047***
		(0.004)	(0.003)	(0.003)		(0.004)
Observations	2,057	2,057	2,057	2,057	2,057	2,057
Student Controls	✓	✓	✓	✓	✓	✓

Notes: This table presents the mediating effects of different mediator combinations on the impact of the algorithmic recommender on recommendation adoption in the *Geography* scenario. Each model iteratively excludes one mediator to assess the sensitivity of the remaining mediator coefficients. Participants rated each mediating channel's associated statement on a scale ranging from -5 to 5. Estimates come from a linear probability model. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S21: Mediated Recommendation Adoption Rates Gap, Pragmatism Scenario

	Outcome: Recommendation Adoption					
	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm	-0.047*	-0.020	-0.015	-0.001	-0.017	-0.015
	(0.023)	(0.026)	(0.024)	(0.023)	(0.023)	(0.024)
Mediators:						
Intent			0.017**	0.028***	0.032***	0.017**
			(0.007)	(0.008)	(0.008)	(0.007)
Ability		0.005		0.004	0.005	0.000
		(0.004)		(0.005)	(0.004)	(0.004)
Comprehension		0.035***	0.030***		0.045***	0.030***
		(0.005)	(0.005)		(0.004)	(0.004)
Alignment		0.050***	0.047***	0.054***		0.047***
		(0.005)	(0.004)	(0.004)		(0.005)
Observations	2,039	2,039	2,039	2,039	2,039	2,039
Student Controls	✓	✓	✓	✓	✓	✓

Notes: This table presents the mediating effects of different mediator combinations on the impact of the algorithmic recommender on recommendation adoption in the *Pragmatism* scenario. Each model iteratively excludes one mediator to assess the sensitivity of the remaining mediator coefficients. Participants rated each mediating channel's associated statement on a scale ranging from -5 to 5. Estimates come from a linear probability model. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. All specifications include school fixed effects. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S22: Cronbach's Alpha Coefficients for Mediator Combinations

	Scenario		
Mediator Combination	Heart	Geography	Pragmatism
Ability, Alignment	0.436	0.297	0.475
Ability, Comprehension	0.443	0.192	0.540
Comprehension, Alignment	0.397	0.328	0.626
Intent, Ability	0.610	0.463	0.647
Intent, Alignment	0.574	0.552	0.612
Intent, Comprehension	0.608	0.500	0.695
Ability, Comprehension, Alignment	0.526	0.366	0.648
Intent, Ability, Alignment	0.639	0.540	0.672
Intent, Ability, Comprehension	0.655	0.492	0.718
Intent, Comprehension, Alignment	0.628	0.564	0.730
Intent, Ability, Comprehension, Alignment	0.679	0.567	0.750

Notes: This table reports Cronbach's alpha coefficients for different combinations of mediators across the three decision-making scenarios.

Table S23: Proportion of Students by Variance Levels in Each Mediator

	Variance		
Mediator	**Low**	**Medium**	**High**
Intent	0.609	0.318	0.072
Ability	0.737	0.225	0.038
Comprehension	0.609	0.298	0.093
Alignment	0.349	0.439	0.212

Notes: This table presents the proportion of students categorized into Low, Medium, and High variance groups for the four mediators. The variance categories were determined based on predefined cutoffs: $Low \leq 3$, $Medium$: 3–10 and $High \geq 10$. The maximum possible variance for this scale (range -5 to 5) is 16.67. A low variance indicates consistent responses across the three scenarios, while a high variance suggests significant fluctuations in responses.

S5 Text Analysis

Table S24: Sentiment Ratings by Recommendation Source

Sentiment	Human	Algorithm	Difference
Positive	0.71 (0.68-0.73)	0.64 (0.62-0.67)	-0.061***
Negative	0.06 (0.05-0.07)	0.09 (0.07-0.11)	0.034***
Neutral	0.19 (0.18-0.21)	0.22 (0.20-0.24)	0.027**

Notes: This table presents the mean sentiment ratings (with 95% confidence intervals in parentheses) of recommendation source sentiment classification. The Difference column represents the mean difference (Algorithm - Human) with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S25: Correlation of Text Topic Scores with Mediators, Heart Scenario

	Mediators			
Topic	Intent	Ability	Comprehension	Alignment
Helpfulness				
Helpful	0.054**	0.143***	0.030	-0.012
Unknown	-0.034	-0.102***	-0.023	0.031
Not Helpful	-0.021	-0.077***	0.024	0.002
Purpose				
Decide for me	0.065***	0.032	0.033	0.018
Direction	0.038	0.063**	0.013	-0.003
Exploration	-0.027	-0.003	-0.016	-0.021
Validation	0.034	0.016	0.044*	0.001
Recommendation Basis				
Career Prospects	0.020	0.032	0.011	-0.030
Grades and Chances	0.044*	0.019	0.003	0.016
Interests	-0.017	0.044*	0.010	-0.096***

Notes: This table reports Pearson correlation coefficients between text classification scores for each topic and agreement scores for each mediator in the *Heart* scenario. Text classification scores are derived from the DeBERTa-v3 transformer model, which is pre-trained and fine-tuned for natural language inference to classify responses and score topic categories independently (Sileo, 2024). Participants rated each mediating channel's associated statement on a scale from -5 to 5. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S26: Correlation of Text Topic Scores with Mediators, Geography Scenario

Topic	Intent	Ability	Comprehension	Alignment
	Mediators			
Helpfulness				
Helpful	0.120***	0.190***	0.084***	0.090***
Unknown	-0.026	-0.072***	-0.088***	-0.018
Not Helpful	-0.063**	-0.106***	-0.062**	-0.047*
Purpose				
Decide for me	0.024	0.004	0.048**	0.004
Direction	0.038	0.094***	0.014	0.028
Exploration	-0.006	0.025	-0.002	0.011
Validation	0.016	0.015	0.065***	-0.019
Recommendation Basis				
Career Prospects	0.029	0.078***	0.028	0.040
Grades and Chances	0.062**	0.026	0.090***	0.057**
Interests	-0.015	0.031	0.033	-0.014

Notes: This table reports Pearson correlation coefficients between text classification scores for each topic and agreement scores for each mediator in the *Geography* scenario. Text classification scores are derived from the DeBERTa-v3 transformer model, which is pre-trained and fine-tuned for natural language inference to classify responses and score topic categories independently (Sileo, 2024). Participants rated each mediating channel's associated statement on a scale from -5 to 5. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S27: Correlation of Text Topic Scores with Mediators, Pragmatism Scenario

	Mediators			
Topic	Intent	Ability	Comprehension	Alignment
Helpfulness				
Helpful	0.134***	0.174***	0.088***	0.097***
Unknown	-0.096***	-0.126***	-0.102***	-0.042*
Not Helpful	-0.099***	-0.112***	-0.064***	-0.056**
Purpose				
Decide for me	0.068***	0.058**	0.083***	0.037
Direction	0.098***	0.091***	0.052**	0.046*
Exploration	0.043*	0.017	0.005	0.049**
Validation	0.052**	0.031	0.074***	-0.008
Recommendation Basis				
Career Prospects	0.071***	0.042*	0.050**	0.074***
Grades and Chances	0.062**	0.054**	0.079***	-0.012
Interests	0.080***	0.079***	0.089***	0.099***

Notes: This table reports Pearson correlation coefficients between text classification scores for each topic and agreement scores for each mediator in the *Pragmatism* scenario. Text classification scores are derived from the DeBERTa-v3 transformer model, which is pre-trained and fine-tuned for natural language inference to classify responses and score topic categories independently (Sileo, 2024). Participants rated each mediating channel's associated statement on a scale from -5 to 5. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S28: Mediation Analysis Using Text Topics

	Helpfulness			Purpose				Recommendation Basis		
Scenario	Helpful	Unknown	Not Helpful	Decide for me	Direction	Exploration	Validation	Career Prospects	Grades and Chances	Interests
Heart										
Indirect Effect	-0.003	-0.003	-0.003**	0.000	-0.002	0.000	0.000	-0.003	0.000	-0.001
	(0.006)	(0.002)	(0.001)	(0.001)	(0.005)	(0.001)	(0.000)	(0.004)	(0.002)	(0.002)
Direct Effect	-0.010	-0.010	-0.011	-0.013	-0.011	-0.013	-0.013	-0.010	-0.013	-0.013
	(0.026)	(0.025)	(0.023)	(0.024)	(0.025)	(0.023)	(0.023)	(0.026)	(0.022)	(0.024)
Topic Score Mean	0.688	0.159	0.024	0.148	0.293	0.283	0.289	0.386	0.293	0.476
Topic Score SD	0.329	0.223	0.095	0.185	0.249	0.256	0.202	0.298	0.249	0.329
Geography										
Indirect Effect	0.002	0.000	-0.003	0.000	0.008	0.000	0.000	0.000	-0.002	0.000
	(0.007)	(0.003)	(0.002)	(0.001)	(0.006)	(0.001)	(0.000)	(0.003)	(0.003)	(0.003)
Direct Effect	-0.039*	-0.037**	-0.034*	-0.037**	-0.045**	-0.037**	-0.037**	-0.036*	-0.035*	-0.037**
	(0.020)	(0.018)	(0.019)	(0.018)	(0.019)	(0.018)	(0.018)	(0.019)	(0.018)	(0.019)
Topic Score Mean	0.688	0.159	0.024	0.148	0.293	0.283	0.289	0.386	0.293	0.476
Topic Score SD	0.329	0.223	0.095	0.185	0.249	0.256	0.202	0.298	0.249	0.329
Pragmatism										
Indirect Effect	-0.002	0.002	-0.002	0.000	0.008*	0.000	0.000	-0.001	0.004*	-0.006***
	(0.009)	(0.003)	(0.003)	(0.001)	(0.005)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Direct Effect	-0.034	-0.038	-0.034	-0.037	-0.045*	-0.037	-0.037	-0.036	-0.041	-0.030
	(0.028)	(0.027)	(0.027)	(0.028)	(0.026)	(0.027)	(0.027)	(0.028)	(0.028)	(0.027)
Topic Score Mean	0.688	0.159	0.024	0.148	0.293	0.283	0.289	0.386	0.293	0.476
Topic Score SD	0.329	0.223	0.095	0.185	0.249	0.256	0.202	0.298	0.249	0.329

Notes: The table reports estimates from linear probability models. Text classification scores are derived from the DeBERTa-v3 transformer model, which is pre-trained and fine-tuned for natural language inference to classify responses and score topic categories independently (Sileo, 2024). All specifications include controls and school fixed effects. Controls include a female indicator, grade level indicators, track indicators, an indicator for intending to participate in the national exams for university admission, norm compliance score tertile indicators, and GPA tertile indicators. To minimize record loss, indicators for missing regressor values were also included. Standard errors clustered at the school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S29: Text Classification Scores for Students Not Inclined to Use Recommender

Topic	Human Counselor	Algorithmic Recommender	Diff
Helpfulness			
Helpful	0.56 (0.50-0.63)	0.47 (0.42-0.52)	-0.098**
Unknown	0.24 (0.20-0.29)	0.25 (0.21-0.29)	0.005
Not Helpful	0.06 (0.03-0.08)	0.08 (0.06-0.11)	0.024
Purpose			
Decide for Me	0.13 (0.10-0.16)	0.11 (0.09-0.13)	-0.022
Direction	0.24 (0.19-0.28)	0.20 (0.18-0.23)	-0.031
Exploration	0.26 (0.22-0.30)	0.25 (0.22-0.28)	-0.007
Validation	0.21 (0.18-0.24)	0.25 (0.23-0.28)	0.042**
Recommendation Basis			
Career Prospects	0.29 (0.24-0.34)	0.29 (0.25-0.33)	-0.001
Grades and Chances	0.20 (0.16-0.24)	0.24 (0.20-0.27)	0.037
Interests	0.44 (0.38-0.50)	0.37 (0.33-0.42)	-0.065*

Notes: This table presents the mean classification scores, with 95% confidence intervals in parentheses, for students who are less inclined to use recommendations. The scores reflect their responses to the question: *"How could such a counselor [as in the scenarios] be more helpful to you?"* across three topics. Students less inclined to use recommendation are those who scored less than zero on a Likert scale ranging from -5 to +5, indicating their (lack of) willingness to use a recommender for college application decisions. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S30: Text Classification Scores by Gender

Category	Human	Algorithm	Difference
Helpfulness			
Helpful			
Male	0.75 (0.71-0.78)	0.66 (0.62-0.71)	-0.086***
Female	0.79 (0.77-0.81)	0.68 (0.65-0.71)	-0.112***
Unknown			
Male	0.12 (0.09-0.14)	0.16 (0.13-0.19)	0.039**
Female	0.12 (0.10-0.13)	0.16 (0.14-0.18)	0.041***
Not Helpful			
Male	0.01 (0.01-0.02)	0.02 (0.01-0.03)	0.005
Female	0.01 (0.00-0.01)	0.02 (0.01-0.03)	0.013***
Purpose			
Decide for me			
Male	0.14 (0.12-0.17)	0.13 (0.11-0.14)	-0.017
Female	0.16 (0.14-0.17)	0.18 (0.16-0.20)	0.019
Direction			
Male	0.30 (0.27-0.34)	0.25 (0.22-0.28)	-0.057**
Female	0.36 (0.34-0.38)	0.29 (0.26-0.31)	-0.073***
Exploration			
Male	0.27 (0.24-0.30)	0.28 (0.25-0.31)	0.009
Female	0.29 (0.27-0.31)	0.31 (0.28-0.34)	0.018
Validation			
Male	0.28 (0.26-0.31)	0.29 (0.26-0.31)	0.002
Female	0.31 (0.30-0.33)	0.30 (0.28-0.32)	-0.008
Recommendation Basis			
Career Prospects			
Male	0.39 (0.35-0.43)	0.37 (0.33-0.41)	-0.020
Female	0.46 (0.43-0.49)	0.38 (0.35-0.40)	-0.084***
Grades and Chances			
Male	0.28 (0.25-0.30)	0.34 (0.30-0.37)	0.059**
Female	0.30 (0.28-0.32)	0.33 (0.31-0.36)	0.029*
Interests			
Male	0.50 (0.46-0.54)	0.45 (0.41-0.49)	-0.046
Female	0.53 (0.50-0.56)	0.48 (0.44-0.51)	-0.055**

Notes: This table presents the text classification scores of *usage*-inclined students segmented by gender. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S31: Text Classification Scores for Helpfulness by GPA Tertiles

Category	Human	Algorithm	Difference
Helpful			
Bottom	0.76 (0.72-0.8)	0.66 (0.61-0.7)	-0.103***
Middle	0.77 (0.74-0.8)	0.66 (0.62-0.71)	-0.11***
Top	0.8 (0.77-0.83)	0.69 (0.65-0.73)	-0.109***
Unknown			
Bottom	0.14 (0.12-0.17)	0.19 (0.16-0.23)	0.05**
Middle	0.12 (0.1-0.15)	0.14 (0.12-0.17)	0.021
Top	0.09 (0.07-0.11)	0.14 (0.11-0.17)	0.052***
Not Helpful			
Bottom	0.01 (0-0.02)	0.02 (0.01-0.04)	0.012*
Middle	0.01 (0-0.01)	0.02 (0.01-0.03)	0.012*
Top	0.01 (0-0.01)	0.01 (0.01-0.02)	0.007*

Notes: This table presents the text classification scores of *usage*-inclined students segmented by GPA tertiles. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S31: (B) Text Classification Scores for Purpose by GPA Tertiles

Category	Human	Algorithm	Difference
Decide for me			
Bottom	0.16 (0.13-0.18)	0.15 (0.13-0.17)	-0.006
Middle	0.15 (0.13-0.18)	0.15 (0.13-0.18)	0
Top	0.15 (0.13-0.18)	0.17 (0.14-0.19)	0.014
Direction			
Bottom	0.36 (0.33-0.4)	0.26 (0.23-0.3)	-0.098***
Middle	0.34 (0.31-0.37)	0.27 (0.24-0.3)	-0.071***
Top	0.32 (0.29-0.35)	0.28 (0.25-0.31)	-0.042**
Exploration			
Bottom	0.28 (0.24-0.31)	0.31 (0.27-0.35)	0.035
Middle	0.29 (0.26-0.32)	0.27 (0.24-0.31)	-0.018
Top	0.29 (0.25-0.32)	0.31 (0.27-0.35)	0.024
Validation			
Bottom	0.32 (0.29-0.34)	0.28 (0.25-0.31)	-0.032*
Middle	0.29 (0.27-0.32)	0.3 (0.27-0.32)	0.003
Top	0.3 (0.28-0.32)	0.3 (0.27-0.33)	0.004

Notes: This table presents the text classification scores of *usage*-inclined students segmented by GPA tertiles. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S31: (C) Text Classification Scores for Recommendation Basis by GPA Tertiles

Category	Human	Algorithm	Difference
Career Prospects			
Bottom	0.42 (0.38-0.46)	0.36 (0.32-0.4)	-0.056*
Middle	0.44 (0.41-0.48)	0.36 (0.32-0.4)	-0.086***
Top	0.45 (0.41-0.49)	0.4 (0.36-0.44)	-0.053*
Grades and Chances			
Bottom	0.29 (0.26-0.32)	0.32 (0.28-0.36)	0.024
Middle	0.27 (0.25-0.3)	0.32 (0.29-0.36)	0.05**
Top	0.32 (0.29-0.34)	0.35 (0.32-0.39)	0.037
Interests			
Bottom	0.51 (0.47-0.55)	0.44 (0.39-0.49)	-0.066**
Middle	0.53 (0.49-0.56)	0.43 (0.39-0.48)	-0.093***
Top	0.53 (0.49-0.57)	0.53 (0.48-0.57)	0.001

Notes: This table presents the text classification scores of *usage*-inclined students segmented by GPA tertiles. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S32: Text Classification Scores by Urbanicity

Category	Human	Algorithm	Difference
Helpfulness			
Helpful			
Rural	0.76 (0.72-0.8)	0.68 (0.64-0.73)	-0.075**
Urban	0.78 (0.76-0.81)	0.67 (0.63-0.7)	-0.117***
Unknown			
Rural	0.14 (0.11-0.16)	0.16 (0.13-0.19)	0.024
Urban	0.11 (0.1-0.13)	0.16 (0.14-0.18)	0.047***
Not Helpful			
Rural	0.01 (0-0.02)	0.02 (0.01-0.03)	0.009
Urban	0.01 (0-0.01)	0.02 (0.01-0.02)	0.011***
Purpose			
Decide for me			
Rural	0.15 (0.12-0.17)	0.16 (0.14-0.19)	0.015
Urban	0.16 (0.14-0.17)	0.16 (0.14-0.17)	0
Direction			
Rural	0.36 (0.32-0.39)	0.29 (0.26-0.33)	-0.063**
Urban	0.33 (0.31-0.36)	0.26 (0.24-0.28)	-0.074***
Exploration			
Rural	0.29 (0.26-0.32)	0.3 (0.26-0.34)	0.01
Urban	0.28 (0.26-0.3)	0.3 (0.27-0.32)	0.015
Validation			
Rural	0.3 (0.27-0.32)	0.31 (0.28-0.34)	0.01
Urban	0.31 (0.29-0.32)	0.29 (0.27-0.31)	-0.013
Recommendation Basis			
Career Prospects			
Rural	0.46 (0.42-0.5)	0.38 (0.35-0.42)	-0.078***
Urban	0.43 (0.4-0.45)	0.37 (0.34-0.4)	-0.058***
Grades and Chances			
Rural	0.31 (0.28-0.34)	0.32 (0.28-0.36)	0.01
Urban	0.29 (0.27-0.31)	0.34 (0.31-0.36)	0.052***
Interests			
Rural	0.5 (0.46-0.54)	0.47 (0.42-0.51)	-0.035
Urban	0.53 (0.5-0.55)	0.47 (0.44-0.5)	-0.061***

Notes: This table presents the text classification scores of *usage*-inclined students segmented by school locale. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S33: (A) Text Classification Scores for Helpfulness by Track

Category	Human	Algorithm	Difference
Helpful			
Humanities	0.78 (0.75-0.82)	0.64 (0.59-0.69)	-0.141***
Life Sciences	0.78 (0.74-0.82)	0.71 (0.66-0.76)	-0.07**
Exact Sciences	0.78 (0.74-0.82)	0.68 (0.62-0.73)	-0.106***
Economics and IT	0.76 (0.72-0.8)	0.67 (0.62-0.72)	-0.094***
Unknown			
Humanities	0.13 (0.10-0.15)	0.16 (0.13-0.20)	0.034
Life Sciences	0.11 (0.08-0.13)	0.16 (0.13-0.2)	0.06***
Exact Sciences	0.11 (0.09-0.14)	0.14 (0.1-0.17)	0.021
Economics and IT	0.13 (0.1-0.15)	0.17 (0.14-0.21)	0.045**
Not Helpful			
Humanities	0.01 (0-0.01)	0.02 (0.01-0.03)	0.015***
Life Sciences	0.01 (0.00-0.01)	0.01 (0.00-0.02)	0.005
Exact Sciences	0.01 (0-0.02)	0.01 (0.01-0.02)	0.001
Economics and IT	0.01 (0-0.02)	0.03 (0.01-0.04)	0.018**

Notes: This table presents the text classification scores of *usage*-inclined students segmented by track. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S33: (B) TEXT CLASSIFICATION SCORES FOR PURPOSE BY TRACK

Category	Human	Algorithm	Difference
Decide for me			
Humanities	0.15 (0.12-0.17)	0.16 (0.13-0.19)	0.014
Life Sciences	0.15 (0.12-0.18)	0.16 (0.13-0.19)	0.015
Exact Sciences	0.18 (0.15-0.21)	0.16 (0.13-0.19)	-0.019
Economics and IT	0.14 (0.11-0.16)	0.15 (0.12-0.17)	0.01
Direction			
Humanities	0.36 (0.33-0.39)	0.28 (0.24-0.32)	-0.078***
Life Sciences	0.34 (0.3-0.38)	0.27 (0.24-0.31)	-0.065**
Exact Sciences	0.31 (0.27-0.34)	0.28 (0.24-0.32)	-0.023
Economics and IT	0.36 (0.32-0.4)	0.25 (0.22-0.28)	-0.107***
Exploration			
Humanities	0.30 (0.26-0.33)	0.26 (0.22-0.29)	-0.042
Life Sciences	0.27 (0.23-0.31)	0.35 (0.30-0.4)	0.08***
Exact Sciences	0.28 (0.24-0.32)	0.26 (0.22-0.29)	-0.021
Economics and IT	0.29 (0.25-0.32)	0.33 (0.29-0.37)	0.044
Validation			
Humanities	0.32 (0.29-0.34)	0.32 (0.29-0.35)	0.003
Life Sciences	0.30 (0.27-0.33)	0.30 (0.26-0.33)	0
Exact Sciences	0.28 (0.26-0.31)	0.30 (0.26-0.33)	0.018
Economics and IT	0.31 (0.28-0.34)	0.27 (0.24-0.3)	-0.041*

Notes: This table presents the text classification scores of *usage*-inclined students segmented by track. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S33: (C) Text Classification Scores for Recommendation Basis by Track

Category	Human	Algorithm	Difference
Career Prospects			
Humanities	0.42 (0.38-0.46)	0.34 (0.3-0.39)	-0.078**
Life Sciences	0.44 (0.4-0.49)	0.39 (0.35-0.44)	-0.047
Exact Sciences	0.46 (0.41-0.5)	0.36 (0.32-0.41)	-0.096***
Economics and IT	0.43 (0.39-0.48)	0.40 (0.35-0.44)	-0.037
Grades and Chances			
Humanities	0.27 (0.24-0.3)	0.33 (0.29-0.37)	0.058**
Life Sciences	0.32 (0.28-0.35)	0.33 (0.29-0.38)	0.016
Exact Sciences	0.31 (0.28-0.34)	0.34 (0.3-0.39)	0.031
Economics and IT	0.28 (0.25-0.31)	0.33 (0.29-0.37)	0.046*
Interests			
Humanities	0.54 (0.5-0.58)	0.45 (0.4-0.5)	-0.095***
Life Sciences	0.49 (0.44-0.54)	0.48 (0.43-0.54)	-0.008
Exact Sciences	0.53 (0.48-0.57)	0.48 (0.43-0.54)	-0.046
Economics and IT	0.51 (0.46-0.55)	0.46 (0.41-0.51)	-0.051

Notes: This table presents the text classification scores of *usage*-inclined students segmented by track. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S34: (A) Text Classification Scores for Helpfulness by Norm Compliance

Category	Human	Algorithm	Difference
Helpful			
Bottom	0.79 (0.75-0.82)	0.64 (0.59-0.69)	-0.148***
Middle	0.80 (0.77-0.83)	0.71 (0.66-0.75)	-0.093***
Top	0.75 (0.71-0.78)	0.67 (0.63-0.71)	-0.079***
Unknown			
Bottom	0.12 (0.10-0.15)	0.15 (0.12-0.19)	0.029
Middle	0.12 (0.10-0.15)	0.13 (0.10-0.16)	0.008
Top	0.11 (0.09-0.13)	0.19 (0.15-0.22)	0.074***
Not Helpful			
Bottom	0.01 (0.00-0.01)	0.02 (0.01-0.03)	0.011**
Middle	0.01 (0.00-0.01)	0.01 (0.00-0.02)	0.004
Top	0.01 (0.00-0.02)	0.03 (0.02-0.04)	0.014**

Notes: This table presents the text classification scores of *usage*-inclined students segmented by Norm Compliance. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S34: (B) Text Classification Scores for Purpose by Norm Compliance

Category	Human	Algorithm	Difference
Decide for me			
Bottom	0.14 (0.11-0.16)	0.16 (0.13-0.18)	0.019
Middle	0.15 (0.12-0.18)	0.16 (0.13-0.18)	0.006
Top	0.17 (0.14-0.19)	0.16 (0.14-0.18)	-0.007
Direction			
Bottom	0.36 (0.32-0.39)	0.26 (0.22-0.29)	-0.103***
Middle	0.34 (0.31-0.38)	0.29 (0.26-0.33)	-0.049**
Top	0.33 (0.3-0.36)	0.26 (0.24-0.29)	-0.061***
Exploration			
Bottom	0.31 (0.27-0.34)	0.3 (0.26-0.34)	-0.009
Middle	0.32 (0.29-0.36)	0.3 (0.26-0.34)	-0.017
Top	0.24 (0.21-0.26)	0.29 (0.26-0.33)	0.055***
Validation			
Bottom	0.29 (0.27-0.32)	0.28 (0.25-0.31)	-0.007
Middle	0.31 (0.28-0.33)	0.31 (0.28-0.34)	0.004
Top	0.31 (0.28-0.33)	0.29 (0.27-0.32)	-0.012

Notes: This table presents the text classification scores of *usage*-inclined students segmented by Norm Compliance. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S34: (C) Text Classification Scores for Recommendation Basis by Norm Compliance

Category	Human	Algorithm	Difference
Career Prospects			
Bottom	0.43 (0.39-0.47)	0.36 (0.32-0.41)	-0.067**
Middle	0.44 (0.40-0.48)	0.39 (0.34-0.43)	-0.058**
Top	0.44 (0.40-0.47)	0.37 (0.34-0.41)	-0.066***
Grades and Chances			
Bottom	0.25 (0.22-0.27)	0.29 (0.25-0.33)	0.042*
Middle	0.30 (0.27-0.33)	0.37 (0.33-0.4)	0.071***
Top	0.32 (0.3-0.35)	0.34 (0.31-0.37)	0.015
Interests			
Bottom	0.55 (0.51-0.59)	0.51 (0.46-0.56)	-0.041
Middle	0.54 (0.5-0.58)	0.47 (0.42-0.51)	-0.069**
Top	0.48 (0.45-0.52)	0.43 (0.39-0.47)	-0.051*

Notes: This table presents the text classification scores of *usage*-inclined students segmented by Norm Compliance. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S35: (A) Text Classification Scores for Helpfulness by Grade

Category	Human	Algorithm	Difference
Helpful			
Grade 10	0.78 (0.75-0.82)	0.67 (0.63-0.72)	-0.109***
Grade 11	0.78 (0.75-0.81)	0.67 (0.63-0.71)	-0.11***
Grade 12	0.76 (0.73-0.8)	0.67 (0.62-0.72)	-0.093***
Unknown			
Grade 10	0.13 (0.11-0.16)	0.18 (0.15-0.22)	0.049**
Grade 11	0.11 (0.09-0.13)	0.16 (0.13-0.19)	0.048***
Grade 12	0.11 (0.08-0.13)	0.12 (0.09-0.15)	0.017
Not Helpful			
Grade 10	0.01 (0-0.02)	0.03 (0.01-0.04)	0.016**
Grade 11	0 (0-0.01)	0.02 (0.01-0.03)	0.013***
Grade 12	0.01 (0-0.02)	0.01 (0.01-0.01)	0.001

Notes: This table presents the text classification scores of *usage*-inclined students segmented by grade. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table S35: (B) Text Classification Scores for Purpose by Grade

Category	Human	Algorithm	Difference
Decide for me			
Grade 10	0.2 (0.17-0.22)	0.17 (0.14-0.19)	-0.027
Grade 11	0.15 (0.12-0.17)	0.15 (0.13-0.17)	0.004
Grade 12	0.11 (0.09-0.13)	0.15 (0.12-0.18)	0.045**
Direction			
Grade 10	0.37 (0.34-0.4)	0.28 (0.24-0.31)	-0.095***
Grade 11	0.33 (0.3-0.36)	0.28 (0.25-0.31)	-0.052**
Grade 12	0.32 (0.28-0.35)	0.25 (0.22-0.29)	-0.062***
Exploration			
Grade 10	0.28 (0.25-0.31)	0.3 (0.26-0.33)	0.019
Grade 11	0.31 (0.27-0.34)	0.31 (0.28-0.35)	0.007
Grade 12	0.27 (0.24-0.3)	0.28 (0.24-0.32)	0.01
Validation			
Grade 10	0.32 (0.3-0.34)	0.3 (0.27-0.32)	-0.025
Grade 11	0.3 (0.27-0.32)	0.29 (0.27-0.32)	-0.005
Grade 12	0.29 (0.27-0.31)	0.31 (0.27-0.34)	0.016

Notes: This table presents the text classification scores of *usage*-inclined students segmented by grade. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

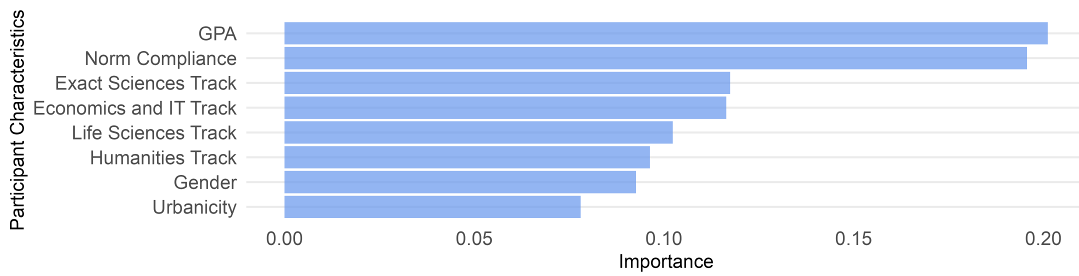
Table S35: (C) Text Classification Scores for Recommendation Basis by Grade

Category	Human	Algorithm	Difference
Career Prospects			
Grade 10	0.46 (0.43-0.5)	0.37 (0.33-0.4)	-0.095***
Grade 11	0.43 (0.39-0.47)	0.38 (0.35-0.42)	-0.047*
Grade 12	0.41 (0.38-0.45)	0.37 (0.32-0.41)	-0.047
Grades and Chances			
Grade 10	0.32 (0.29-0.35)	0.36 (0.32-0.4)	0.041*
Grade 11	0.31 (0.29-0.34)	0.33 (0.3-0.37)	0.017
Grade 12	0.24 (0.21-0.27)	0.3 (0.26-0.34)	0.059**
Interests			
Grade 10	0.54 (0.5-0.57)	0.46 (0.42-0.5)	-0.078***
Grade 11	0.49 (0.45-0.53)	0.47 (0.43-0.51)	-0.023
Grade 12	0.53 (0.49-0.57)	0.48 (0.43-0.53)	-0.056*

Notes: This table presents the text classification scores of *usage*-inclined students segmented by grade. The Difference column represents the mean difference (Algorithm - Human), with * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

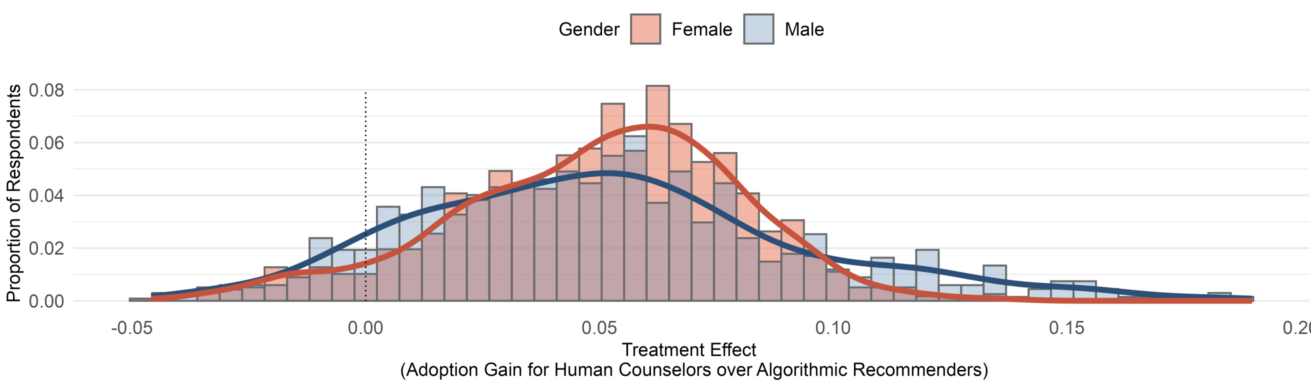
S6 Targeted Treatment Assignment

Figure S3: Variable Importance from Fitted Causal Forest



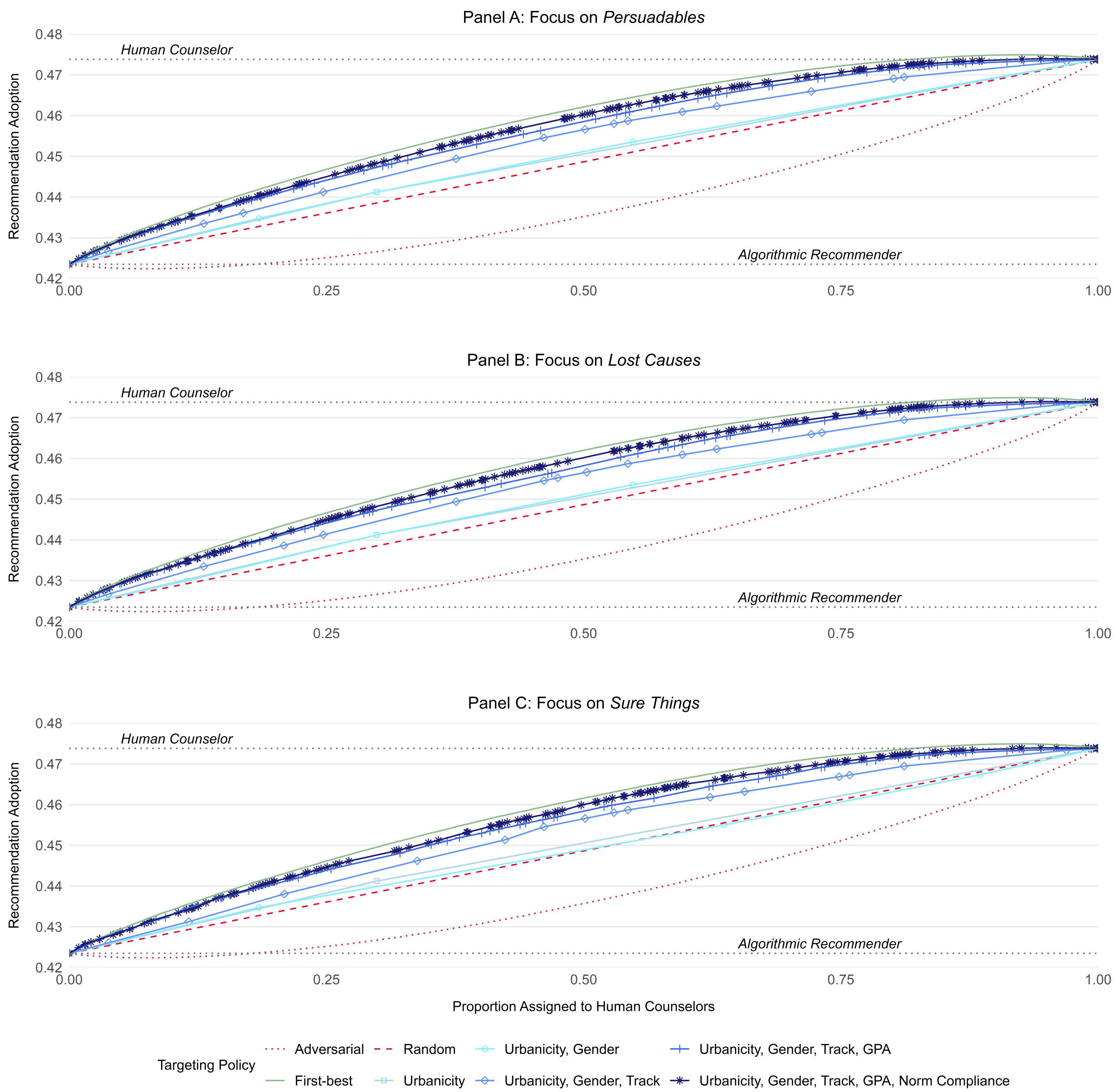
Notes: The variable importance plot displays the relative contribution of each participant characteristics to the fitted Causal Forest model. Variables with higher importance scores have a greater influence on estimating heterogeneous treatment effects within our dataset.

Figure S4: Heterogeneity in Treatment Effect based on Gender



Notes: This figure shows the frequency distribution of treatment effect estimates derived from the Causal Forest grouped by gender. The spread of the histogram reflects the heterogeneity in treatment effects of female and male students, indicating that variation in adoption increases when a human counselor is provided compared to an algorithmic recommender.

Figure S5: Central Planner's Welfare Under Different Prioritization for Human Counselor Assignment



Notes: This figure plots the central planner's welfare function as an increasing share of the targeted population that is assigned to human counselors, under three different preference assumptions. In the first panel (*Persuadables*), the planner aims to maximize adoption rate improvements by treating all baseline adoption levels equally in terms of preference for adoption rate increments ($g(p) = p$). In the second panel (*Lost Causes*), planners prioritize adoption rate improvements among individuals with lower baseline adoption rates ($g(p) = p^{0.5}$). In the third panel (*Sure Things*), planners prioritize adoption rate improvements among individuals with higher baseline adoption rates ($g(p) = p^2$).

Table S36: Subgroup Prioritization for Counselor Assignment Assuming Focus on Persuadables

Order	Urbanicity	Gender	Track	GPA	Share	Adoption Rate
1	Urban	Male	Exact Sciences	Low	0.0292	0.4271
2	Rural	Male	Exact Sciences	Low	0.0492	0.4292
3	Rural	Male	Life Sciences	Low	0.0508	0.4294
4	Urban	Male	Exact Sciences	Medium	0.0735	0.4313
5	Rural	Female	Humanities	High	0.0854	0.4323
6	Rural	Female	Humanities	Medium	0.1059	0.4340
7	Rural	Male	Exact Sciences	Medium	0.1129	0.4345
8	Urban	Female	Exact Sciences	Low	0.1361	0.4363
9	Rural	Female	Life Sciences	Low	0.1469	0.4371

Notes: The table shows the top nine subgroups that the planner should prioritize for human counselor assignment. We assume planners prioritize maximizing adoption rate improvements (i.e, equal preference for adoption rate increments at any baseline (algorithmic) adoption). We also assume access to information regarding school urbanicity, student gender, track choice and tertile of students' prior-year GPA. Table 9 presents results when access to student attribute information is more limited (no information on prior-year GPA). With more detailed information on student characteristics, the human counselor assignment process can be more refined. The table also reports the cumulative share in our sample and the estimated cumulative adoption rate under the human counselor condition. To assign human counselors to all nine groups shown, the planner will require human counselors for no more than 15% of the population.

Table S37: Subgroup Prioritization for Counselor Assignment Assuming Focus on Lost Causes

Order	Urbanicity	Gender	Track	GPA	Share	Adoption Rate
1	Urban	Male	Exact Sciences	Low	0.0292	0.4271
2	Rural	Male	Exact Sciences	Low	0.0492	0.4292
3	Rural	Male	Life Sciences	Low	0.0508	0.4294
4	Urban	Male	Exact Sciences	Medium	0.0735	0.4313
5	Rural	Female	Humanities	High	0.0854	0.4323
6	Rural	Male	Exact Science	Medium	0.0924	0.4328
7	Rural	Female	Humanities	Medium	0.1129	0.4345
8	Urban	Female	Exact Sciences	Low	0.1361	0.4363

Notes: The table shows the top subgroups that the planner should prioritize for human counselor assignment. We assume planners prioritize adoption rate improvements among groups with lower baseline (algorithmic) adoption. We also assume access to information regarding school urbanicity, student gender, track choice and tertile of students' prior-year GPA. Table 9 presents results when access to student attribute information is more limited (no information on prior-year GPA). With more detailed information on student characteristics, the human counselor assignment process can be more refined. The table also reports the cumulative share in our sample and the estimated cumulative adoption rate under the human counselor condition. To assign human counselors to all nine groups shown, the planner will require human counselors for no more than 14% of the population.

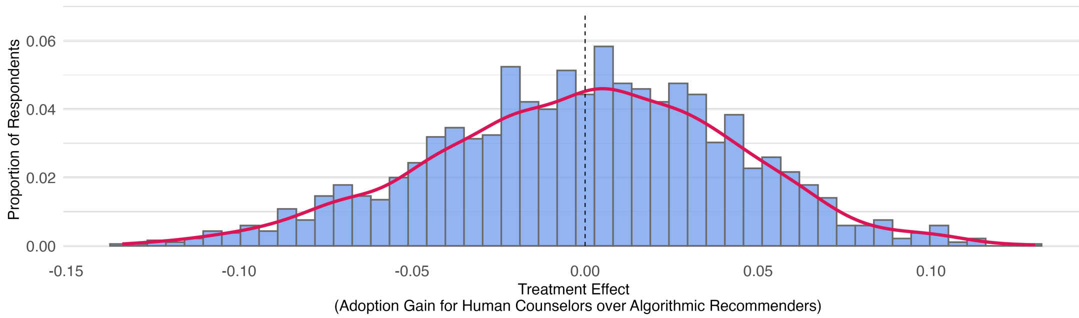
Table S38: Subgroup Prioritization for Counselor Assignment Assuming Focus on Sure Things

Order	**Urbanicity**	**Gender**	**Track**	**GPA**	**Share**	**Adoption Rate**
1	Urban	Male	Exact Sciences	Low	0.0292	0.4271
2	Rural	Female	Humanities	Mid	0.0497	0.4288
3	Rural	Female	Humanities	High	0.0616	0.4297
4	Urban	Male	Exact Sciences	Mid	0.0843	0.4317
5	Rural	Female	Life Sciences	Mid	0.0935	0.4323
6	Rural	Male	Exact Sciences	Low	0.1135	0.4345

Notes: The table shows the top subgroups that the planner should prioritize for human counselor assignment. We assume planners prioritize adoption rate improvements among groups with higher baseline (algorithmic) adoption. We also assume access to information regarding school urbanicity, student gender, track choice and tertile of students' prior-year GPA. Table 9 presents results when access to student attribute information is more limited (no information on prior-year GPA). With more detailed information on student characteristics, the human counselor assignment process can be more refined. The table also reports the cumulative share in our sample and the estimated cumulative adoption rate under the human counselor condition. To assign human counselors to all nine groups shown, the planner will require human counselors for no more than 12% of the population.

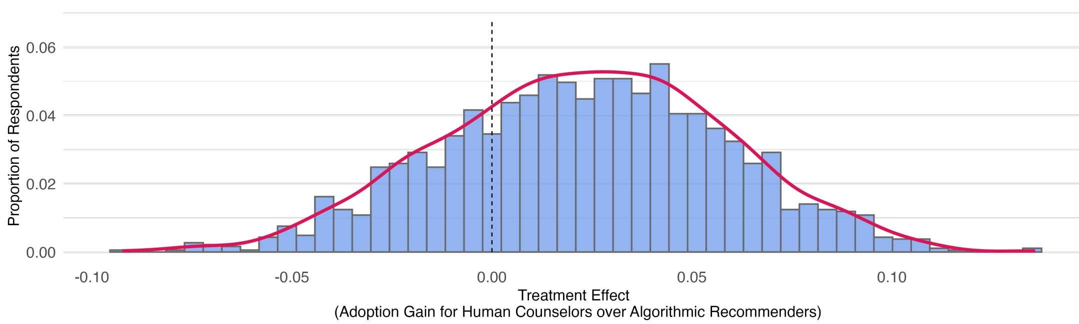
S7 Targeted Policy for Other Scenarios

Figure S6: CATE Distribution for the Heart Scenario



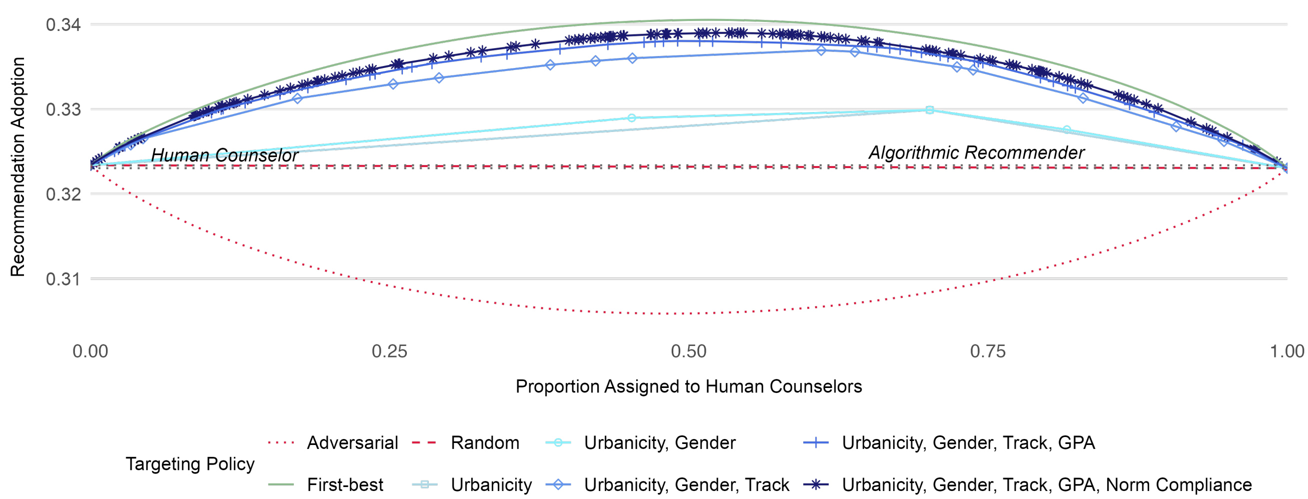
Notes: This figure shows the frequency distribution of treatment effect estimates derived from the Causal Forest for the *Heart* scenario. The spread of the histogram reflects the heterogeneity in treatment effects, indicating variations in adoption gains when a human counselor is provided compared to an algorithmic recommender.

Figure S7: CATE Distribution for the Geography Scenario



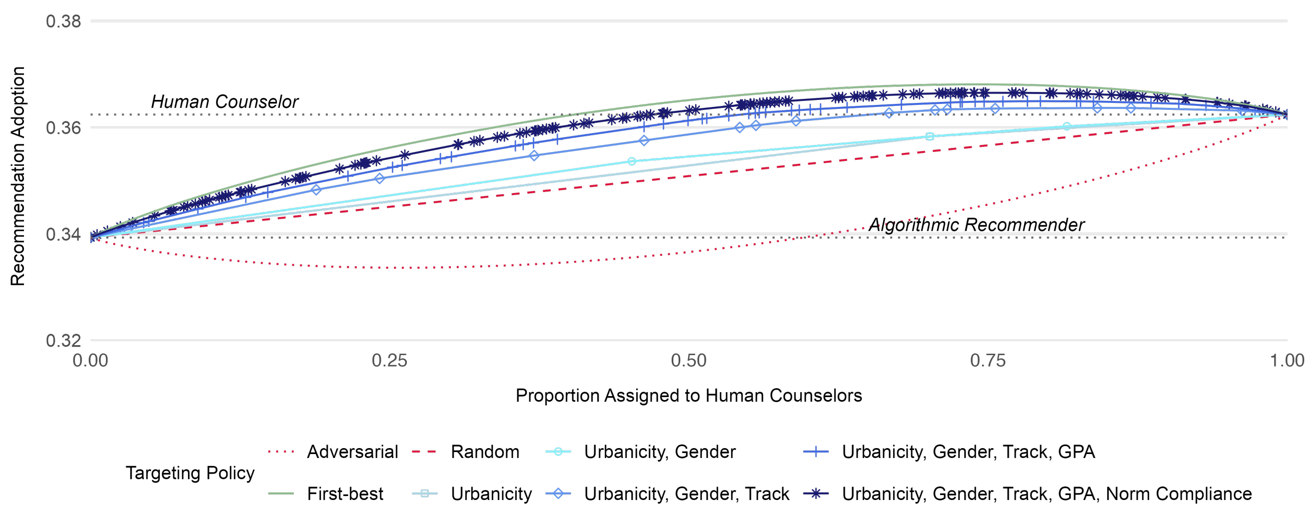
Notes: This figure shows the frequency distribution of treatment effect estimates derived from the Causal Forest for the *Geography* scenario. The spread of the histogram reflects the heterogeneity in treatment effects, indicating variations in adoption gains when a human counselor is provided compared to an algorithmic recommender.

Figure S8: Targeting Policy Performance for the Heart Scenario



Notes: This figure shows the overall adoption rate under different targeting policies as the share of the targeted population increases. For these results, we assume the planner focuses on *Persuadables* (i.e., $g(p) = p$). The area under each curve (AUC) represents the overall adoption rate when the entire population is targeted. The performance of a targeting policy is measured by the difference between the AUC under a specific policy and the AUC under a baseline policy. (In Table 8, we consider either the adversarial policy or the random policy as the baseline.) The *first-best* targeting policy prioritizes the allocation of human counselors to individuals with the highest estimated treatment effects, representing the best possible allocation policy in terms of efficiency. Conversely, the *adversarial* targeting policy prioritizes the allocation of human counselors to individuals with the lowest estimated treatment effects, representing the worst possible allocation policy in terms of efficiency. Both the first-best and adversarial targeting policies involve individual-level targeting based on estimated individual treatment effects. The *random* targeting policy allocates human counselors uniformly at random, independent of the student's characteristics. The remaining strategies involve subgroup-level targeting, utilizing information about school locale (urbanicity), student gender, track choice, prior-year Grade Point Average (GPA), and norm compliance.

Figure S9: Targeting Policy Performance for the Geography Scenario



Notes: This figure shows the overall adoption rate under different targeting policies as the share of the targeted population increases. For these results, we assume the planner focuses on *Persuadables* (i.e., $g(p) = p$). The area under each curve (AUC) represents the overall adoption rate when the entire population is targeted. The performance of a targeting policy is measured by the difference between the AUC under a specific policy and the AUC under a baseline policy. (In Table 8, we consider either the adversarial policy or the random policy as the baseline.) The *first-best* targeting policy prioritizes the allocation of human counselors to individuals with the highest estimated treatment effects, representing the best possible allocation policy in terms of efficiency. Conversely, the *adversarial* targeting policy prioritizes the allocation of human counselors to individuals with the lowest estimated treatment effects, representing the worst possible allocation policy in terms of efficiency. Both the first-best and adversarial targeting policies involve individual-level targeting based on estimated individual treatment effects. The *random* targeting policy allocates human counselors uniformly at random, independent of the student's characteristics. The remaining strategies involve subgroup-level targeting, utilizing information about school locale (urbanicity), student gender, track choice, prior-year Grade Point Average (GPA), and norm compliance.

S8 Alternate Approach for Targeted Treatment Assignment

In the alternate approach, we employed a three-step approach to evaluate treatment effects and optimize treatment assignment. First, we utilized a Causal Forest model incorporating all covariates to estimate individualized treatment outcomes. This method allowed us to leverage the robustness of Causal Forests in capturing heterogeneous treatment effects. Second, using the treatment effect estimates derived from the Causal Forest, we optimized human counselor allocation to maximize recommendation adoption (focus on *Persuadables*. This step ensures that the treatment is allocated to individuals most likely to benefit, improving the overall adoption of the recommendation. Finally, we employed a decision tree algorithm to identify subgroups with distinct treatment response patterns. By stratifying the population into interpretable subgroups, provides population subgroups for targeted assignment of human counselors.

The approach here differs from the main analysis primarily in the sequence of steps. In the alternate approach, treatment outcomes are first determined for the entire population using a Causal Forest, followed by treatment assignment optimization and then subgroup identification based on decision trees. In contrast, the main analysis starts with defining subgroups, then estimates treatment outcomes for these predefined subgroups using the Causal Forest, and finally optimizes treatment assignment. This distinction highlights that the main analysis prioritizes subgroup-specific insights upfront, while the supplementary analysis emphasizes individual treatment effect estimation before stratification.

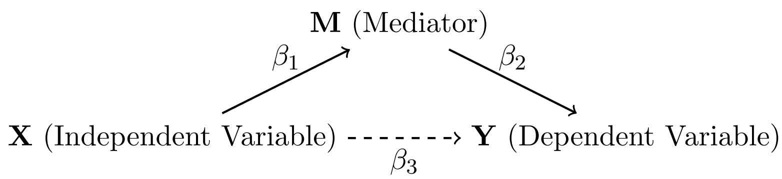
Table S39: Subgroup Prioritization for Counselor Assignment, Alternative Optimization Procedure

Focus on *Persuadables*					
Order	Gender	Track	GPA	Urbanicity	Norm Compliance
1	-	Exact Sciences	<16	-	-
2	Female	Humanities	>18	Rural	-
3	-	Exact Sciences	16 to 18	-	>0.62
4	-	Life Sciences	<16	-	<-1.4
5	-	Economics and IT	<17	Urban	<-1.4
Focus on *Lost Causes*					
Order	Gender	Track	GPA	Urbanicity	Norm Compliance
1	-	Exact Sciences	<16	-	-
2	Male	Exact Sciences	16 to 18	-	>0.5
3	Female	Exact Sciences	16 to 18	-	-
4	-	Economics and IT, Life Sciences	<17	-	<-1.4
Focus on *Sure Things*					
Order	Gender	Track	GPA	Urbanicity	Norm Compliance
1	Male	Life Sciences, Exact Sciences	17 to 18	-	<-4.7
2	Female	Humanities	>17	Rural	-
3	-	Exact Sciences	<17	-	-
4	Female	Life Sciences, Exact Sciences	15 to 17	-	<-1.7

Notes: This table presents results from an alternative optimization procedure that employs a decision tree to identify target student subgroups based on estimated outcomes from a causal forest, incorporating all available school and student attributes to optimize treatment assignment (Athey and Wager, 2019, 2021). In the first scenario (*Persuadables*), the planner aims to maximize adoption rate improvements by treating all baseline adoption levels equally in terms of preference for adoption rate increments ($g(p) = p$). In the second scenario (*Lost Causes*), planners prioritize adoption rate improvements among individuals with lower baseline adoption rates ($g(p) = p^{0.5}$). In the third scenario (*Sure Things*), planners prioritize adoption rate improvements among individuals with higher baseline adoption rates ($g(p) = p^2$).

S9 Explainer: Mediation Analysis

Mediation analysis is a statistical approach used to understand *how* and *why* an independent variable (X) influences a dependent variable (Y) through an *intermediate variable*, called a **mediator** (M). This helps determine if an effect is direct or if it operates through another variable.



A typical mediation model consists of the following regression equations:

1. Effect of X on M (Path a)

$$M = \beta_1 X + \epsilon_1 \tag{3}$$

where β_1 captures the effect of X on the mediator M.

2. Effect of M on Y (Path b), controlling for X

$$Y = \beta_2 M + \beta_3 X + \epsilon_2 \tag{4}$$

where:

- β_2 represents the effect of the mediator M on the outcome Y.
- β_3 captures any remaining direct effect of X on Y.

3. Total Effect of X on Y (Path c)

$$Y = \beta_4 X + \epsilon_3 \tag{5}$$

where β_4 represents the **total effect** of X on Y before accounting for mediation.

Key Interpretations

- The **direct effect** of X on Y is given by β_3.
- The **indirect (mediated) effect** is computed as the product of a and b:

$$\text{Indirect Effect} = \beta_1 \beta_2 \tag{6}$$

- The **total effect** is the sum of direct and indirect effects:

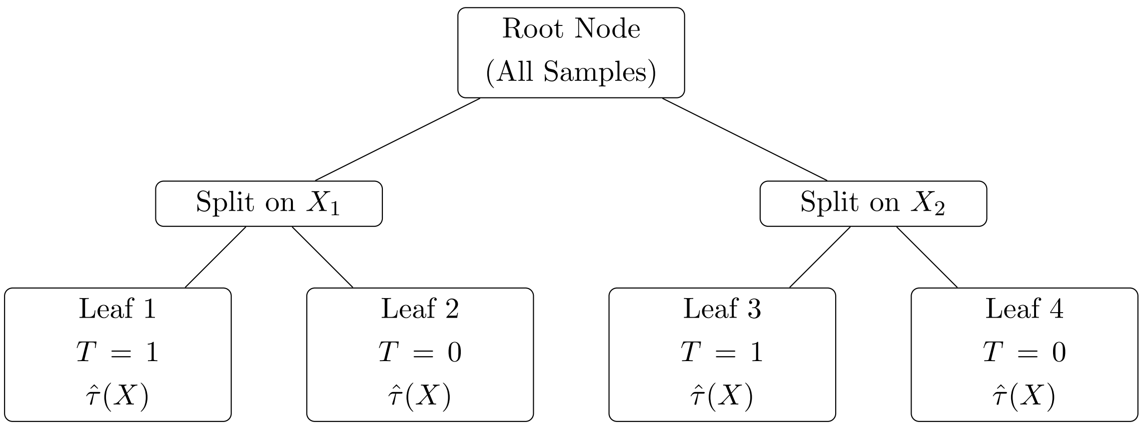
$$\text{Total Effect} = \beta_3 + (\beta_1 \beta_2) \tag{7}$$

In our context, we have results for path a in Table S17, path b in Tables S19 - S21, and path c in Table 2. The Mediation Analysis in Table 5 is a culmination of the three of these.

S10 Explainer: Causal Forest

Causal Forests are a machine learning method designed for estimating heterogeneous treatment effects in observational and experimental studies. Unlike traditional regression models, Causal Forests use a tree-based structure to estimate treatment effects at an individual or subgroup level, making them particularly useful for Conditional Average Treatment Effect (CATE) estimation.

Visual Representation of a Tree from a Causal Forest



Explanation:

- The tree starts from a **root node**, which considers all samples.
- It **splits** based on covariates $(X_1, X_2, X_3, \ldots)$.
- The final nodes (**leaves**) contain treatment effect estimates $\hat{\tau}(X)$.
- Multiple trees are aggregated in the causal forest to obtain the final treatment effect estimate.

Key Estimation Equations

1. Individual Treatment Effect

$$\hat{\tau}(X) = \mathbb{E}[Y|X, T = 1] - \mathbb{E}[Y|X, T = 0] \tag{8}$$

2. Random Forest Aggregation

$$\hat{\tau}(X) = \frac{1}{B} \sum_{b=1}^{B} \hat{\tau}_b(X) \tag{9}$$

where B is the number of trees in the forest.