

OrientAI

A French AI-driven orientation platform: conversational, sourced, and supervised for post-secondary and career choices.

Application file
AI Grand Challenge 2026

Presented by

Matteo Lepietre

Ella Faouzi

Paris Dauphine-PSL University

Organized by INRIA

May 11, 2026

Table of contents

Executive Summary	4
1. Introduction	5
2. Literature Review	6
2.1. Post-baccalaureate reorientation: scale and margins for action	6
2.2. Structural inequalities in access to guidance	7
2.3. Generative AI as a new risk factor	8
3. Quantitative Pilot: Survey on Students' Guidance Experience	9
3.1. Description of the sample	10
3.2. Mapping of the declared guidance experience	10
3.2.1. Guidance that relies on the private circle more than on the public service 11	
3.2.2. Overall perceptions: a moderate sentiment, polarised at the margins	12
3.3. Quality of counselling and reorientation: a link more complex than it appears 13	
3.4. Social origin and guidance trajectory: three convergent mechanisms of inequality	14
3.4.1. Parental education level and the quality of counselling received: a non-significant tendency	14
3.4.2. The chosen track reflects social origin	15
3.4.3. Some tracks concentrate reorientations	17
3.4.4. A transversal reading: three mechanisms that reinforce one another	17
3.5. Use of generative AI and perceived information	18
3.5.1. A use already widespread, but unequal across contexts	18
3.5.2. The use of generative AI does not improve the sense of being informed ..	19
3.5.3. A reading of the observed non-effect	19
3.6. Synthesis and transversal reading of the pilot	20
3.6.1. Three central results	20
3.6.2. The moment of reorientation: a peak in L3 to be interpreted with caution	20
3.6.3. Overall limitations of the pilot	21
4. Our Solution: The OrientAI Platform	22
4.1. A Sunday evening	22
4.2. A three-module architecture	22
4.2.1. The AI guidance assistant	22
4.2.2. The human relay	23

4.2.3.	<i>Temporal continuity</i>	24
4.3.	<i>Accessibility</i>	25
5.	<i>OrientAI: architecture of the conversational module and benchmark</i>	25
5.1.	<i>Why a system grounded in French sources rather than a generalist LLM</i>	26
5.2.	<i>Pipeline overview</i>	26
5.3.	<i>The corpus: 47,214 records from 25 official public sources</i>	28
5.4.	<i>Searching within the corpus: vector indexing and hybrid retrieval</i>	28
5.5.	<i>Framed generation: the strict contract imposed on the LLM</i>	29
5.6.	<i>Anti-hallucination safeguards: a multi-layered apparatus</i>	31
5.7.	<i>Initial benchmark: comparison with generalist LLMs</i>	32
5.8.	<i>Data governance and regulatory compliance</i>	33
5.9.	<i>Acknowledged limitations and conditions for national scaling</i>	34
6.	<i>Implementation: institutional anchoring, deployment, governance</i>	35
6.1.	<i>Institutional and economic anchoring within the AVENIR(s) programme</i>	36
6.2.	<i>The human conditions of deployment</i>	38
6.3.	<i>The conditions for governance over time</i>	39
7.	<i>Conclusion</i>	40
8.	<i>References</i>	41
8.1.	<i>Career guidance, education, and public policy</i>	41
8.2.	<i>AI, algorithmic bias, and ethical issues in education</i>	43
8.3.	<i>RAG architecture and LLM techniques</i>	43
8.4.	<i>Appendix</i>	44

Executive Summary

Navigating French higher education has become a cognitively demanding undertaking, in a landscape where the public guidance service struggles to fill the informational deficit experienced by young people. For pupils least supported by the institution, the transmission of strategic information relies largely on family cultural capital, a socially and geographically unequal resource. Within this vacant space, a new source of information has established itself in only a few years: generative AI, now used by virtually all French students surveyed. This massive adoption precedes any public reflection on two nonetheless major issues: the sovereignty of the particularly sensitive data that young people entrust to these tools, and the documented structural biases of generalist language models, biases that are at once commercial, geographic, socio-economic, and demographic, which silently shape their recommendations.

The OrientAI project proposes a response to this twofold deficit: a public platform of guidance support, designed for the French context and compliant by construction with European requirements regarding high-risk AI systems. It is articulated around three complementary modules. A conversational assistant grounded in a RAG architecture anchored in official public sources, which answers questions without social judgement, without unsourced figures, and without prescriptive recommendations. A connection to state-recognised guidance counsellors, who take over whenever a situation calls for a human dimension that AI cannot sustain. Finally, a dashboard centralising the calendars and administrative procedures currently scattered across several public portals.

The present report sets out two preliminary empirical explorations: a quantitative pilot that seeks to document at the individual level phenomena that the French literature primarily addresses in aggregate, and a benchmark comparing a technical prototype with generalist LLMs on a handful of representative guidance cases. It also develops the conditions for implementation, with an operational anchoring envisaged within the AVENIR(s) programme, the public infrastructure already deployed across all French *collèges* and *lycées*. Conducted by two students drawing on their own experience of the system, this work brings to the debate a perspective that is, as of today, largely absent from it. We assume its status as a prototype and sketch the work that an eventual move to a larger scale would call for.

1. Introduction

Navigating French higher education has become a cognitively demanding undertaking. More than 25,000 post-secondary programmes coexist on Parcoursup, within a landscape where the boundaries between public and private, selective and non-selective, vocational and academic tracks have blurred. Faced with this complexity, more than one young person in two now considers that they lacked information to make their choice, and nearly one in three reorient themselves after entering higher education (see section 2). This difficulty in choosing is not an individual problem: it is the symptom of an information system whose fragmentation is no longer commensurate with the decisions it is meant to inform.

The public guidance service struggles to fill this gap. National Education psychologists, whose numbers have fallen by nearly a third over forty years, are now structurally directed by the institution itself towards the most vulnerable populations. For other students, the transmission of strategic information relies largely on family cultural capital, a socially and geographically unequal resource. In this vacant space, a new source of information has established itself within only a few years: generative AI, now used by virtually all French students surveyed, and which directly influences the orientation choices of 52% of them.

This massive adoption precedes by several years any public reflection on two nonetheless major issues. The first concerns sovereignty: the particularly sensitive data that young people entrust to these tools, including academic results, aspirations, and family circumstances, transit to foreign operators outside any national educational framework, with no oversight of their subsequent use or of the models that ingest them. The second concerns the documented structural biases of language models. A rapidly expanding international empirical literature establishes that these biases are simultaneously commercial, geographic, socio-economic, and demographic. They operate silently, without the user's awareness, and steer recommendations according to rationales unrelated to the relevance of the advice for the person seeking it.

Our project proposes to address this twofold observation through a public guidance support platform, designed for the French context and compliant by design with European requirements regarding high-risk AI systems. The platform brings together three complementary modules. A conversational assistant powered by OrientAI, a RAG architecture built on Mistral and grounded in authoritative public sources, which answers users' questions about programmes, professions, and career outcomes without social judgement, without unsourced figures, and without prescriptive recommendations. A connection with state-recognised guidance counsellors (National Education psychologists, university SCUIO services, Mission Locale advisors), who take over whenever a situation calls for a human dimension that AI cannot sustain. Finally, a dashboard that centralises the calendars and administrative procedures currently scattered across more than six official platforms (Parcoursup, MonMaster, CROUS, CVEC, CAF, messervices.etudiant.gouv.fr), covering the entirety of a nine-month trajectory.

Three choices structure our approach. First, the scope: the platform addresses the entire guidance trajectory, from final-year upper-secondary students to adults in career transition, rather than a single targeted audience. Second, the stance: we conceive of OrientAI as an augmented information service, not as a substitute for human counselling, which remains the responsibility of guidance counsellors and on which the platform concentrates available time for the conversations that genuinely require it. Finally, a two-stage empirical approach: a quantitative pilot study aiming to document at the individual level phenomena that the French

literature primarily addresses in aggregate, and a technical prototype whose preliminary results are compared with those of generalist LLMs on French guidance cases.

This report presents the two empirical contributions and outlines the platform project they underpin. Section 2 reviews the literature that illuminates the guidance issues to which our project responds. Section 3 presents the methodology and results of the quantitative pilot. Section 4 describes the OrientAI platform and its three modules through a representative usage session. Section 5 sets out the conversational prototype, its architecture, its benchmark protocol, and its initial results. Section 6 develops the conditions for implementing the project: institutional anchoring, human conditions of deployment, and long-term governance. Section 7 concludes.

2. Literature Review

Post-baccalaureate guidance is now a well-documented research object, at the crossroads of economics, sociology, and more recently studies on algorithmic bias. The review that follows traces what this literature establishes about the causes of reorientation, the inequalities in access to counselling, and the new role played by generative AI in young people's information practices.

2.1. Post-baccalaureate reorientation: scale and margins for action

Post-baccalaureate reorientation has become a phenomenon of significant scale within French higher education. IPP Note no. 122, published in January 2026 by the economist Nagui Bechichi and based on the exhaustive administrative data of APB, Parcoursup, SISE, and SCOLARITÉ, establishes that in 2022, one third of upper-secondary leavers who had just entered higher education submitted new applications on Parcoursup, compared with 21% ten years earlier under Admission Post-Bac (Bechichi, 2026). The note is unambiguous on one essential point: reorientation is not a failure, but a corrective mechanism. This is precisely what makes it a target for optimisation: if reorientation restores trajectories downstream, then preventing initial misallocations, as the author advocates, would reduce both educational costs and the number of interrupted trajectories. Bechichi estimates the gross additional cost per narrowly reoriented student at €13,200.

Not all motives for reorientation, however, can be addressed by better guidance. Bechichi mentions several: the reaction to a rejection from one's preferred application, the reassessment of an initial choice owing to a misjudgement of a programme's content and its fit with the student's profile, or the wish to rebound after a situation close to academic failure. Among these mechanisms, only that of a poorly informed initial choice stems from a guidance deficit and constitutes the direct target of our study. The others, whether non-admission to a preferred application, academic difficulty, or an evolution of one's personal project during studies, fall by nature outside the reach of any upstream improvement in counselling.

It is precisely this mechanism that the thematic report of the General Inspectorate of Education, Sport and Research identifies as a priority. The report notes a continuous increase in reorientations affecting all types of students, regardless of their academic profile or track, and cites, drawing on data from the Cour des comptes, a 24% growth between 2017 and 2019 in the number of reorientation candidates who accepted an offer on Parcoursup; more broadly,

133,000 reorientation candidates had registered on the platform in 2019 (IGÉSR, 2020). Above all, the mission proposes an operational response that prefigures OrientAI's rationale: enriching Parcoursup programme descriptions with concrete information on course content, study plans, and campus life, so as to enable candidates to assess the fit between a programme and their own profile before applying. The institutional diagnosis and our proposal converge on this observation: the information currently available remains too abstract to enable an informed choice.

This informational deficit is largely confirmed by young people's own perceptions. The OpinionWay survey conducted in April 2025 for Edumapper on a representative sample of 1,064 young people aged 18 to 24 excluding upper-secondary students (quota method, ISO 20252) establishes that 56% regret their guidance choice owing to a lack of information, and that only 38% consider that pupils are well prepared to choose their professional path (OpinionWay & Edumapper, 2025). The priority expectations expressed are precise: 31% call for more testimonies from students and professionals, and 29% for easier access to information on career outcomes. Two features that current public resources document without making them directly accessible in a conversational logic adapted to each profile.

2.2. Structural inequalities in access to guidance

The informational deficit documented in 2.1 does not affect French young people uniformly. It is distributed along lines of social and geographic fracture that turn the quality of the counselling received into an inequality variable in its own right. Two phenomena fuel these inequalities: public guidance counselling is saturated, and families compensate for this shortfall in highly unequal ways depending on their social background.

The number of National Education psychologists (PsyEN), the only professionals statutorily responsible for guidance counselling in secondary education, has fallen sharply over forty years. According to the data drawn upon in a written question tabled in the National Assembly in December 2025, approximately 3,330 PsyEN were assigned to secondary education in 2024, compared with nearly 4,700 guidance counsellors (the predecessors of the current corps) in the early 1980s, even though upper-secondary enrolments have risen (Assemblée nationale, 2025). In 2024, the IGÉSR published a report specifically devoted to the guidance specialism of PsyEN. Seven years after the creation of the corps in 2017, the assessment is severe: the mission explicitly recommends refocusing their activity on supporting populations facing particular difficulties of schooling, learning, guidance, or distress, and advocates strengthening guidance services with pedagogical expertise through the creation of a certification in guidance and the establishment of guidance leads in schools (IGÉSR, 2024). In other words, the public institution itself acknowledges that the PsyEN can no longer, structurally, provide universal guidance counselling: the mission must be redeployed towards the most vulnerable populations, which mechanically leaves other pupils dependent on external informational resources.

Yet these external resources are highly unequal. The 2025 OpinionWay-Edumapper survey quantifies the gap in parental support according to the father's qualifications: 83% of young people benefit from their father's support when he holds a two-year post-secondary qualification, compared with only 68% when he holds no baccalaureate (Ekole, 2025). This quantitative gap of fifteen points in parental support conceals a deeper qualitative gap regarding the very nature of the help provided. A parent with a higher education qualification does not merely offer support: they transmit strategic information that is difficult to acquire otherwise,

such as how to distinguish between similar bachelor's degree titles (MIASHS, MIAGE, computer science), how to compare the actual selectivity of two engineering schools, or how to identify high-quality programmes with limited visibility in search engines. Bourdieu and Passeron, in *Les Héritiers* (1964), had theorised this mechanism: family cultural capital functions as informational capital transmitted silently, enabling children from advantaged backgrounds to make better-calibrated choices without even perceiving this resource as an advantage.

IPP Note no. 93 by Camille Terrier, Rustamdjan Hakimov, and Renke Schmacker (July 2023) provides a recent empirical demonstration of this mechanism. Drawing on a survey of 2,034 final-year upper-secondary students, matched with Parcoursup administrative data, the authors establish that among the top-performing pupils, girls and pupils from disadvantaged social backgrounds have markedly less self-confidence than boys and pupils from advantaged social backgrounds, and that self-confidence is strongly associated with the selectivity of the programmes to which pupils apply (Terrier et al., 2023). The result is important for our study: at identical academic record, some pupils apply to less ambitious programmes because they underestimate their relative position. The inequality therefore concerns not only access to information about programmes; it also concerns the information available to the pupil about themselves and about their objective chances. The authors show that a simple informational intervention, consisting of indicating to each candidate their position within the grade distribution, significantly alters the applications submitted, which demonstrates experimentally that better informational calibration produces more ambitious choices among pupils who initially underestimated their position.

This dual inequality, social and informational, calls for a response complementing the human counselling that, as the IGÉSR confirms, must refocus on the most vulnerable populations.

2.3. Generative AI as a new risk factor

The irruption of generative AI into student practices has been massive and rapid. The *Born AI 2025* report by the Heaven agency, based on an online survey of 495 young people aged 18 to 25 with quotas applied for age, gender, and location, establishes that 93% of respondents used at least one AI-powered service over the past six months, an eight-point increase compared with 2024 (Heaven, 2025). The joint study by Pôle Léonard de Vinci and Talan, covering 1,119 French students in their fourth year of higher education, goes even further: 99% of students use generative AI, including 92% regularly and one third daily, and crucially 52% acknowledge that AI influences their choices (Pôle Léonard de Vinci & Talan, 2024). This last figure is particularly significant for our study: it indicates that AI is no longer merely a productivity tool, but a guidance actor whose influence is now directly acknowledged by students themselves.

Generalist language models, by their very architecture, are not neutral in the recommendations they produce. The foundational work of Bender, Gebru, McMillan-Major, and Mitchell (2021) established the principle: LLMs reproduce and amplify the biases of the corpora on which they are trained, the composition of which massively overrepresents English-language and US-origin content (Bender et al., 2021). This general finding has since been refined by a series of empirical audits focusing specifically on academic recommendations.

The work of Shailya, Mishra, Krishnan, and Ravindran (Centre for Responsible AI, IIT Madras, 2025) tests the university recommendations of several LLMs (LLaMA-3.1-8B, Gemma-7B,

Mistral-7B) using synthetic student profiles. Their results converge: institutions from the Global North are disproportionately favoured, the recommendations reinforce gender stereotypes, and an effect of institutional repetition is prevalent (Shailya et al., 2025). A parallel study by Gupta and Ranjan (Gupta & Ranjan, 2024), comparing the personas generated by GPT-3.5, Gemini, and Claude 3 Sonnet with real LinkedIn data, shows that the three models significantly overweight Stanford, MIT, Berkeley, and Harvard in the generation of profiles of tech professionals. More troubling still, an intersectional audit published in 2026, covering four models (GPT-4o, GPT-4o-mini, Qwen 2.5, GPT-OSS), shows that LLMs adjust the quality of the pedagogical content they produce according to students' demographic attributes, independently of their actual abilities. Family income is the attribute most systematically used by the models as an implicit indicator of intellectual level: at identical academic profile, a student presented as coming from a modest background receives explanations of lower complexity than those provided to a student presented as coming from an advantaged background. The bias persists even when the modest-background student attends an elite institution (Gupta et al., 2026).

It remains to be seen to what extent these biases, documented on Anglophone and international cases, transpose to the French landscape. To our knowledge, no French study published to date has audited the recommendations of ChatGPT, Gemini, or Claude on post-baccalaureate guidance questions, which constitutes precisely one of the objectives of the benchmark presented in section 5.

The literature presented in fact describes three facets of a single mechanism: a guidance choice is built on information whose quantity, quality, and source vary considerably from one student to another. Economic work on reorientation documents the aggregate outcome of these variations without being able to isolate their individual determinants. Research in the sociology of education and in experimental economics establishes that these variations are not random but socially structured. Recent audits of LLMs add a new variable whose effect on French young people remains unknown. These three dimensions have never been jointly explored on the same individuals within the French context. Such is the ambition, necessarily modest at this stage, of the pilot presented in section 3.

3. Quantitative Pilot: Survey on Students' Guidance Experience

Between April and May 2026, 180 students and recently active young adults, primarily from the ecosystem of Université Paris-Dauphine, responded to a questionnaire concerning their actual guidance experience: the resources on which they relied to make their choices, their retrospective assessment of the advice received, their use of generative AI throughout this trajectory, and the links between these dimensions and their social origin or their reorientation paths. The survey is exploratory and does not claim national representativeness; it aims to establish initial empirical elements making it possible to frame the design of the OrientAI service and to identify the dimensions of inequality that it will need to address as a priority.

The analysis is structured in six sub-sections: description of the sample (3.1), mapping of the declared guidance experience (3.2), then the three levels of the analysis plan: quality of counselling and reorientation (3.3), social determinants of guidance (3.4), use of generative AI (3.5), before a transversal reading and a framing of the overall limitations (3.6).

3.1. Description of the sample

The quantitative pilot relies on a questionnaire disseminated between April and May 2026 through several channels: personal and professional social networks (Facebook, LinkedIn, WhatsApp), institutional emails, and the SurveyCircle platform. As we are both enrolled at Université Paris-Dauphine, dissemination through our personal networks mechanically oriented the composition of the sample towards this ecosystem. In total, 180 responses were collected and analysed after cleaning. The questionnaire primarily targeted students at L2 level and beyond in order to ensure a minimum of hindsight on the post-baccalaureate guidance trajectory. Seven first-year respondents (3.9%) were also retained and are distinguished from the others in the analyses where this is warranted; four of them had already undergone a reorientation at the time of the survey, and their experience documents the share of trajectories that are reconfigured as early as the first year of higher education. The sample also includes six young people recently exited from the educational system (3.3%). Although they were no longer students at the time of the questionnaire, these respondents had a comparable post-baccalaureate guidance trajectory and were retained for this reason.

The profile of the respondents presents three striking characteristics that should be framed before any interpretation. First, the sample is socially advantaged: 60% of respondents declare that at least one of their parents holds a five-year post-secondary qualification or higher, and 84% have at least one parent with a higher education qualification. The category of least-qualified parents (CAP/BEP or no qualification) accounts for only 11% of the sample, that is, 20 respondents. Second, the academic level of the respondents is high: 94.4% obtained a general baccalaureate, and 73% obtained a distinction (*mention Bien* or *Très bien*). Third, the distribution by track reveals a concentration in economics, management and business (55%), followed by humanities and social sciences (13.3%), science and mathematics (12.8%), and law (8.3%).

From a geographic standpoint, 44.4% of respondents completed their secondary education in a medium-sized town, 30% in a large city or in Paris, 17.2% in a small town, 6.1% abroad, and 2.2% in a rural municipality. The distribution by gender reveals a marked over-representation of women (65.6% women against 33.9% men and 0.6% non-binary respondents).

This composition mechanically limits the generalisable scope of the pilot: the results presented in the following sections cannot be extended to French higher education as a whole. They describe the guidance experiences of an academically and socially advantaged audience. Furthermore, the respondents' current institution was not asked for in the questionnaire; given the dissemination channels, a substantial share of the sample is likely to come from Dauphine, although we cannot quantify this precisely. This social and academic homogeneity nevertheless lends particular weight to the social inequalities that the data bring to light (cf. section 3.3): if such inequalities are observed within so homogeneous a sample, their magnitude in the general student population is probably more marked still.

3.2. Mapping of the declared guidance experience

Before examining the links between the variables (sections 3.3 to 3.5), it is appropriate to map the guidance experience as the respondents describe it, distinguishing the sources that guided their initial choice and their retrospective perceptions of the quality of the support received.

3.2.1. Guidance that relies on the private circle more than on the public service

When asked what mainly motivated their initial choice of programme (Q10), respondents massively designate their private circle. Family, immediate circle, and personal network constitute the principal source of motivation for 53.6% of respondents, that is, more than half of the sample. Next come, at substantially lower levels, open days and fairs (12.2%), independent Internet research outside AI (11.2%), and the absence of any structured source, a choice made at random or by default, which concerns 11.7% of respondents. At the opposite end of the spectrum, the institutional actors of guidance counselling carry marginal weight: only 2.2% of respondents cite the guidance counsellor (PsyEN) as the leading factor that guided their choice, and 1.1% mention artificial intelligence. The recoding of open-ended responses also brought to light two categories that the form's initial options did not capture: upper-secondary teachers (3.9% of respondents), and personal motivation or passion (3.4%).

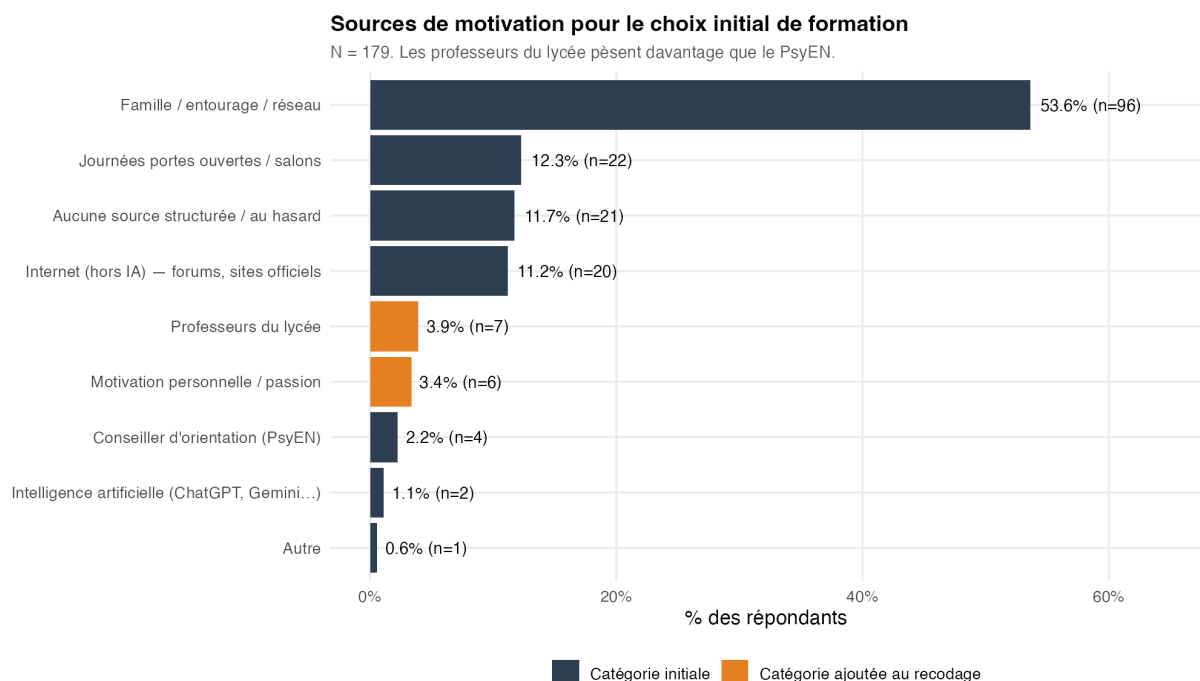


Figure 1 – Sources of motivation for the initial choice of programme. Responses to question Q10 "What mainly motivated your initial choices of programme?". OrientAI pilot, N = 179.

This result calls for a reading that goes beyond mere description. The fact that upper-secondary teachers weigh almost twice as much as the PsyEN as a principal source of motivation directly raises the question of the place currently occupied by the public guidance service in pupils' trajectories. Teachers, who are not trained in guidance and for whom this is not the primary mission, are mobilised more than the professionals dedicated to this task. This inversion of roles is confirmed by the declared frequency of contact with a guidance counsellor in upper-secondary school: 55% of respondents never met a PsyEN during their secondary studies, 31.1% met one only once, and only 13.9% twice or more. The PsyEN, theoretically the pivot of guidance counselling in upper-secondary school, is thus a near-absent actor in pupils' actual trajectories, including within a socially advantaged sample that should, in principle, have better access to these services. This absence effectively leaves guidance to rest on families' private resources, which constitutes fertile ground for the social inequalities examined in section 3.4.

3.2.2. Overall perceptions: a moderate sentiment, polarised at the margins

The respondents also evaluated three dimensions of their guidance experience on a scale of 1 (very poor) to 5 (very good): their sense of being informed about the chosen programme before entering it (Q13), the fit of the advice received with their profile and aspirations (Q14), and their sense of preparedness for entering the labour market at the end of their current programme (Q15). The distributions obtained are centred on close median values, around 3 out of 5 ($\mu = 3.14$ for Q13; $\mu = 3.25$ for Q14; $\mu = 3.27$ for Q15), suggesting at first glance a lukewarm rather than markedly negative or positive sentiment.

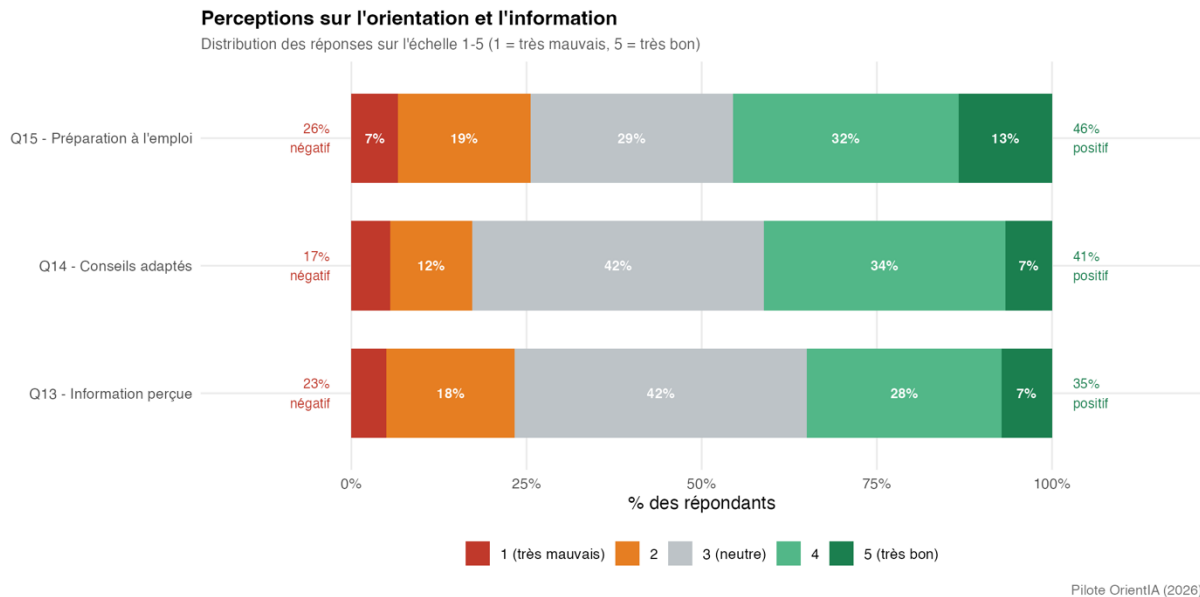


Figure 2 – Distribution of perceptions on guidance and information. Responses to questions Q13 (information perceived before the choice of programme), Q14 (fit of advice received with profile), and Q15 (preparedness for entering the labour market), on a scale of 1 (very poor) to 5 (very good). OrientAI pilot, $N = 180$.

This initial reading by means conceals, however, significant distributions at the margins. Concerning the information perceived before the choice of programme (Q13), 23% of respondents report a negative sense of being informed (scores of 1 or 2) against 35% a positive one (scores of 4 or 5). The gap is more favourable concerning the quality of the advice received (Q14), with 17% negative evaluations against 41% positive, but this still means that one in six respondents retrospectively considers having received poor advice. On preparedness for entering the labour market (Q15), 26% of respondents declare themselves poorly prepared, the highest rate of dissatisfaction across the three dimensions. While the dominant tone remains neutral or moderately positive, the non-negligible share of dissatisfied respondents, particularly on information before the choice and on preparedness for employment, confirms the existence of an audience for whom the current support does not fully fulfil its function.

3.3. Quality of counselling and reorientation: a link more complex than it appears

The first level of the analysis plan investigates the link between the quality of the advice received in upper-secondary school (Q14, scale 1-5) and the probability of reorientation in higher education (Q16). The intuitive hypothesis associates "poor counselling" with "poor choice leading to reorientation". Our data deliver a more nuanced result.

At the bivariate level, we observe no statistically significant difference between reoriented and non-reoriented respondents on the quality of the advice received ($\mu = 3.40$ against $\mu = 3.19$; Mann-Whitney: $p = 0.231$), nor on the sense of being informed before the choice of programme ($\mu = 3.21$ against $\mu = 3.11$; $p = 0.823$). To confirm this apparent absence of a link, we constructed a logistic regression model seeking to predict the probability of reorientation from these two variables. Compared with a reference model without any predictor, the addition of counselling and information does not significantly improve prediction ($p = 0.680$), and only explains a negligible share of the observed variability (pseudo- $R^2 < 1\%$). The pilot therefore does not enable us to bring out a direct effect of the quality of counselling on actual reorientation.

This apparently counter-intuitive result must be read alongside a complementary datum from the same questionnaire. When the 53 reoriented respondents are asked directly whether better advice could have prevented their reorientation (Q16.3), the distribution of responses is strongly polarised: 56.6% consider that better advice would have changed nothing (scores of 1 or 2), while 30.2% consider on the contrary that it would probably have prevented their reorientation (scores of 4 or 5). Nearly a third of the reoriented thus explicitly attribute their trajectory to a counselling deficit in upper-secondary school, something that a global correlation cannot capture.

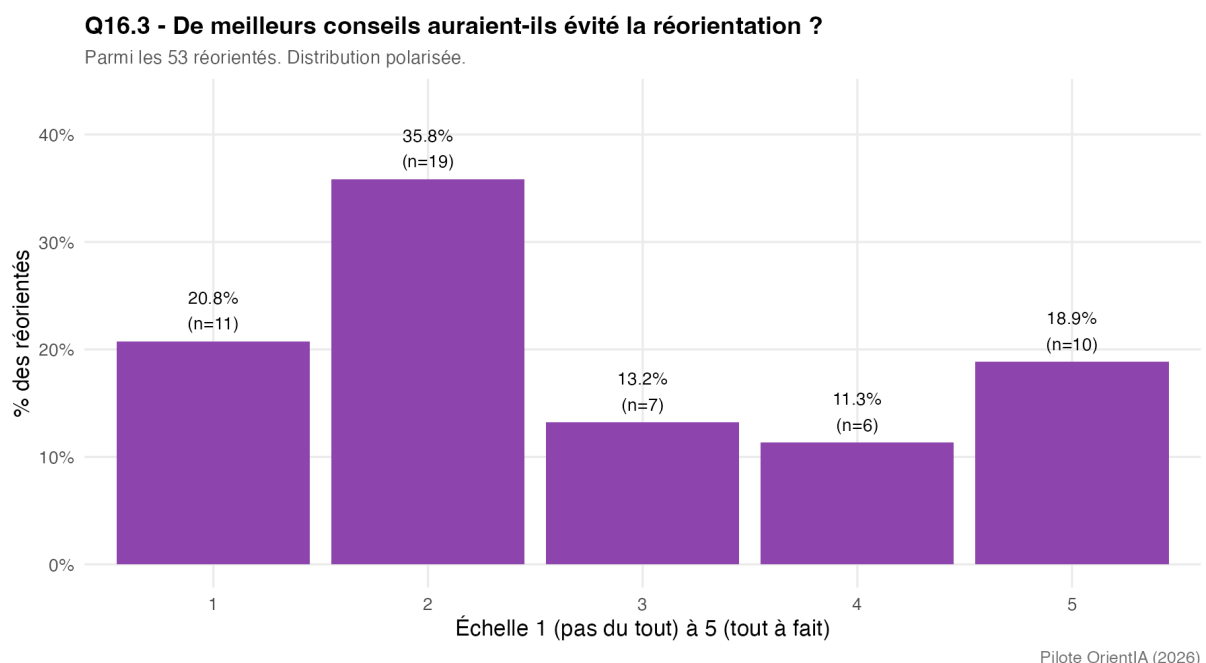


Figure 7 – Distribution of responses to question Q16.3: "Could better advice have prevented your reorientation?". Sample restricted to respondents having declared a reorientation. OrientAI pilot, $N = 53$.

Examining the declared motives for reorientation (Q16.2) sheds light on this duality. Of the 53 reoriented respondents, 26 cite a misalignment with their deeper aspirations (49%), 4 cite poor knowledge of the programme before entering it (7.5%), 7 cite academic difficulty (13%), 2 a personal or family reason (4%), and 14 tick "Other" (26%). This diversity of motives structures the interpretation that follows. Some pertain to dynamics that upper-secondary counselling cannot, by construction, prevent: academic difficulties encountered in higher education, like personal evolutions arising during studies, largely escape the scope of action of pre-baccalaureate guidance. Other motives, by contrast, grouping poor knowledge of the programme and misalignment with aspirations, may refer back to a deficit of information or of guided introspection at the moment of the initial choice. These two categories concern close to 57% of the reoriented respondents in our sample.

This observation circumscribes the scope of the OrientAI service. It could not aim at the elimination of all reorientations, a portion of which results from evolutions of the personal project belonging to the normal dynamics of the student trajectory. By contrast, more than half of the reorientations observed refer to motives that structured support could potentially address upstream.

Two methodological limitations accompany these results. The retrospective evaluation of the counselling received may introduce a desirability bias, with respondents reconstructing their narrative in a direction favourable or unfavourable to upper-secondary school according to their trajectory. Furthermore, the concentration of Q14 evaluations on the median scores (83% at 3, 4, or 5) limits the statistical power to detect an effect specific to poor-quality counselling, which is under-represented in our sample.

3.4. Social origin and guidance trajectory: three convergent mechanisms of inequality

The analysis in section 3.3 showed that the quality of the counselling received had no directly detectable effect on the probability of reorientation. This section proposes a complementary reading by interrogating the social determinants of guidance, through three convergent results: two statistically significant and one descriptive tendency consistent with the literature, which, taken separately, each document one dimension of inequality, and which, articulated together, sketch a coherent pattern.

3.4.1. Parental education level and the quality of counselling received: a non-significant tendency

The parental education level appears associated with the perceived quality of the counselling received in upper-secondary school (Q14) in a manner consistent with a hypothesis of social inequality. Across the five interpretable categories of parental qualification, a rising tendency emerges: $\mu = 2.92$ for parents holding a CAP/BEP, $\mu = 3.00$ for the Baccalauréat, $\mu = 3.07$ for Bac+2 or Bac+3, and $\mu = 3.37$ for Bac+5 and beyond. The children of the most qualified parents thus evaluate the quality of the guidance advice received more favourably than those from less qualified families. This tendency does not, however, reach the threshold of statistical significance on our sample (Kruskal-Wallis on the five interpretable categories: $H = 6.38$; $df = 4$; $p = 0.173$), which invites a cautious reading.

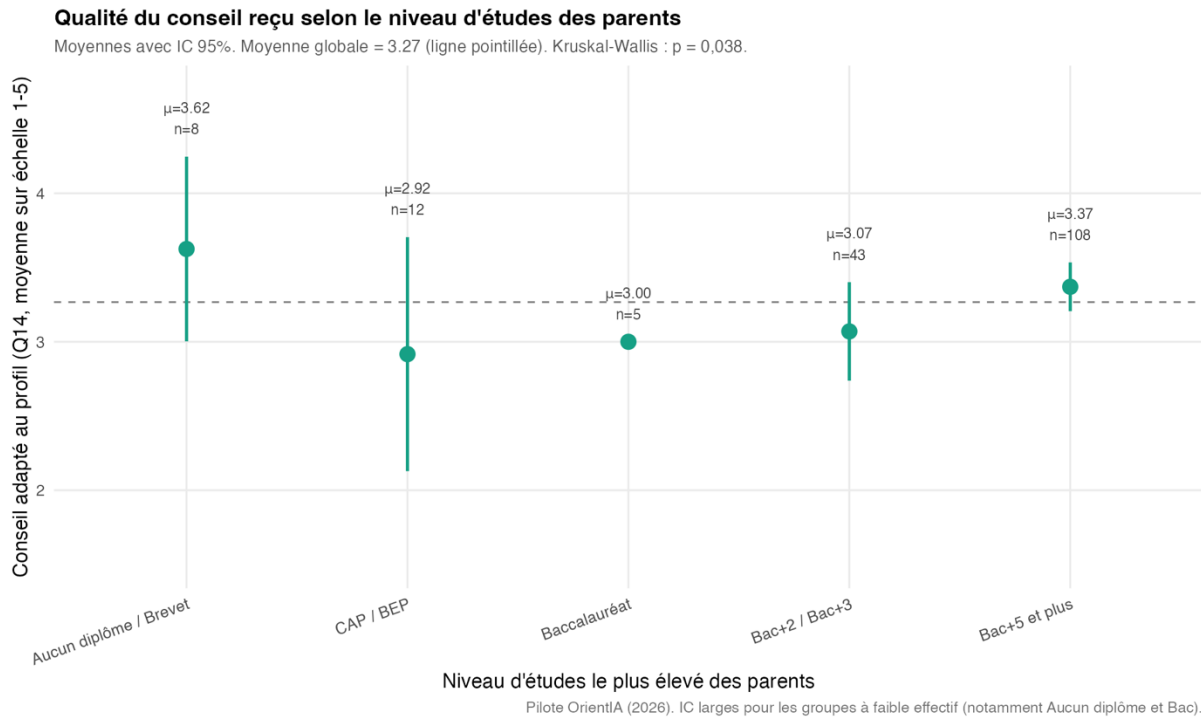


Figure 3 – Quality of the counselling received (Q14) according to the highest education level of the parents. Means with 95% confidence intervals. Kruskal-Wallis test on the five interpretable categories of qualification: $H = 6.38$; $p = 0.173$ (non-significant). OrientAI pilot, $N = 176$.

An apparent exception concerns the "No qualification / Brevet" category, for which the mean ($\mu = 3.62$) exceeds that of the other groups. This anomaly rests, however, on a small subsample ($n = 8$) and a wide confidence interval, which invites caution in its interpretation. On the groups with more substantial subsamples, the tendency is coherent: as the parental education level rises, the quality of the counselling received also progresses. Without reaching statistical significance on our sample, this tendency suggests a first dimension of social inequality whose direction is consistent with the literature (cf. section 2.2): the current system's capacity to deliver adapted counselling does not appear to be uniformly distributed according to pupils' social origin.

3.4.2. The chosen track reflects social origin

The second result concerns the link between the parental education level and the track actually chosen in higher education. For this test, the tracks were regrouped into three coherent pedagogical categories: selective schools (business, engineering, art, health), university tracks in economics or law, and generalist university tracks (humanities, social sciences, sciences). The Chi-square test brings out a significant link between social origin and chosen track ($p = 0.033$, confirmed by a Fisher's exact test under Monte Carlo simulation: $p = 0.036$).

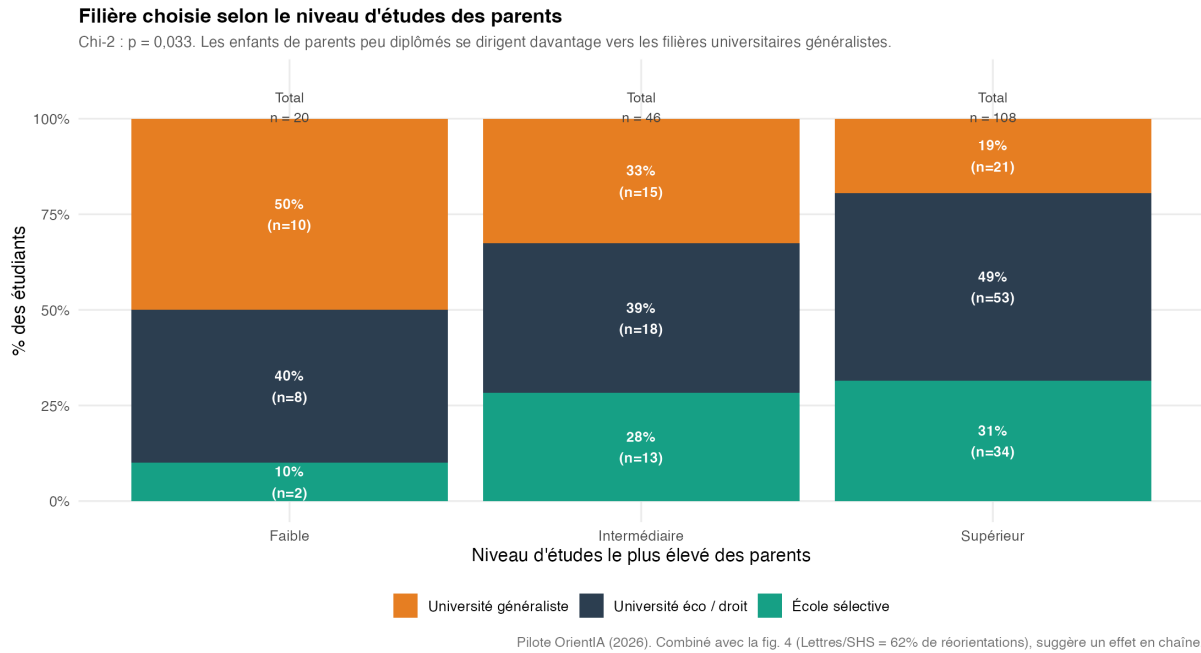


Figure 8 – Chosen track according to parental education level, regrouped into three categories. Chi-square test: $p = 0.033$. OrientAI pilot, $N = 174$.

To analyse the link between social origin and chosen track, the eight categories of tracks in the questionnaire were regrouped into three coherent sets defined as follows. Selective schools include business and management schools, engineering schools, art, design, and architecture schools, as well as health tracks (medicine, pharmacy, physiotherapy, nursing). The criterion retained is the mode of admission: these institutions recruit by national competitive examination, by school-specific competition, or by competitive file review, and operate with limited intakes. University tracks in economics and law correspond to university programmes in economics, management, business, and in law and political science. These tracks are delivered principally at university, without a national entrance examination, but with selectivity that varies across institutions (some universities select on file, others do not). Generalist university tracks regroup programmes in humanities and social sciences on the one hand, and in sciences, mathematics, and computer science on the other. The criterion retained is the absence of national selection at entry and a primary organisation within the public university.

One limitation should be highlighted. Université Paris-Dauphine, from which an important share of our respondents come, is a public university but selective on file, which places it in an intermediate position between the first two categories. To preserve a homogeneous classification (university vs school), we attached it to the "university economics/law" category. This choice may attenuate the contrast measured between the groups, but it ensures consistency with the nomenclature of the French higher education system.

The distribution across groups is eloquent. Among respondents whose parents have a low education level (CAP/BEP or no qualification), 50% are enrolled in a generalist university track, against only 19% of children of parents holding a Bac+5 or above. Conversely, selective schools host 31% of students from highly qualified families against only 10% of students from less qualified families, a ratio of more than three to one. This result documents a second dimension of inequality: at equivalent academic trajectory, our sample being composed of students already admitted to higher education, the track actually chosen reflects social origin.

3.4.3. Some tracks concentrate reorientations

The third result concerns the link between the initially chosen track and the probability of subsequent reorientation. The Chi-square test brings out a strong dependence between these two variables ($p = 0.006$, confirmed by a Fisher's exact test under Monte Carlo simulation: $p = 0.010$), confirming that the reorientation rate varies substantially according to the original track.

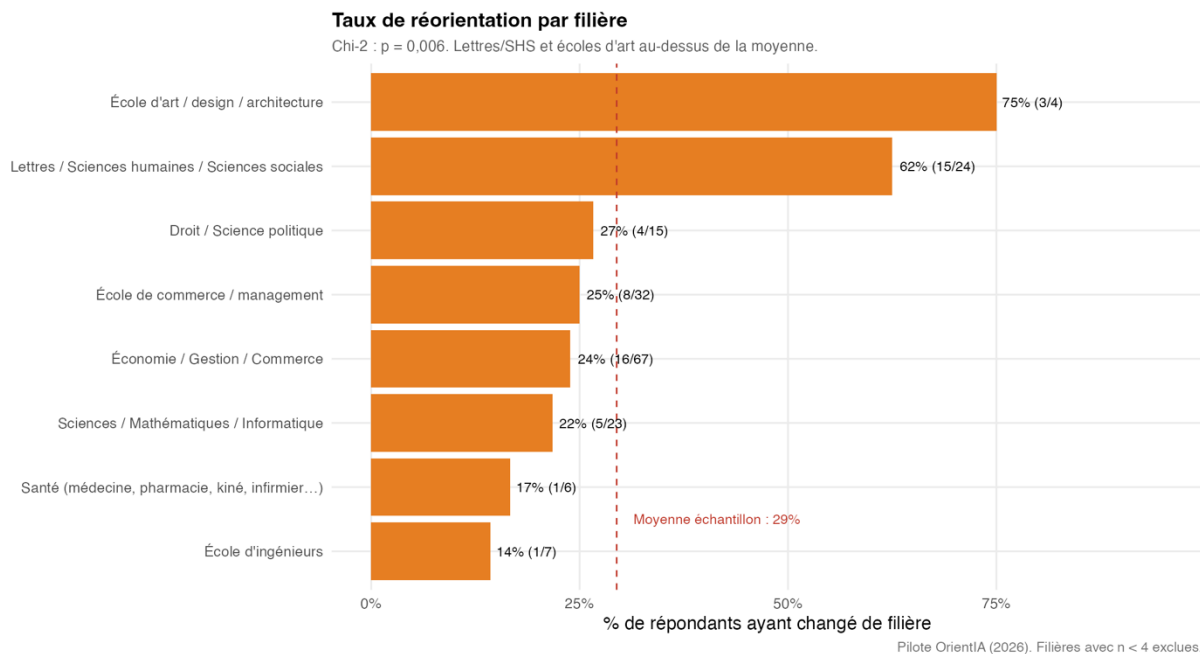


Figure 4 – Reorientation rate by initial track. Chi-square test: $p = 0.006$. OrientAI pilot, $N = 178$.

The observed hierarchy is marked. Generalist university tracks display the highest reorientation rates: 62% of respondents initially enrolled in Humanities / Social Sciences reoriented (15 out of 24), and 75% of those enrolled in art schools (3 out of 4, a subsample to be interpreted with caution). At the opposite end, selective tracks present the lowest rates: 14% in engineering schools (1 out of 7) and 17% in health (1 out of 6). University tracks in economics and law fall into an intermediate zone, at 24% and 27% respectively, close to the sample mean (29%). These gaps are substantial and cannot be attributed to random effects on so clear-cut subsamples. They suggest that the risk of having to change pathway in higher education depends very strongly on the track initially chosen.

3.4.4. A transversal reading: three mechanisms that reinforce one another

Taken separately, each of these three results documents one dimension of inequality in the guidance trajectory. Articulated together, they sketch a coherent pattern of mechanisms that mutually reinforce one another. Children from less qualified families tend to report counselling less adapted to their profile (3.4.1, non-significant tendency). They orient themselves more towards generalist university tracks (3.4.2, $p = 0.033$). These tracks moreover concentrate the highest reorientation rates (3.4.3, $p = 0.006$). While our data do not enable us to formally test a causal chain, owing to the limited size of the sample, the cluster of converging elements, two significant and one consistent in tendency with the literature, is sufficiently clear to be

underlined: inequality of access to quality counselling appears to translate into inequality of placement in higher education, which itself appears to translate into inequality in the stability of the student trajectory. These convergent results plead in favour of an accessible, structured public guidance service, independent of families' private informational resources, which is precisely the objective of the OrientAI service developed in section 5.

3.5. Use of generative AI and perceived information

The third level of the analysis plan interrogates the respondents' current use of generative AI in their guidance trajectory, and examines whether this use improves their sense of being informed about programmes and professions. This analysis is of particular importance in the context of the OrientAI project, which precisely proposes a guidance service supported by a specialised AI.

3.5.1. A use already widespread, but unequal across contexts

The use of generalist AI such as ChatGPT or Gemini is now present in guidance trajectories, but at very different levels according to the objective pursued. For the choice of programme (Q11), 11.1% of respondents declare having mobilised a generative AI, of which 10% as a complement to other steps and only 1.1% exclusively. For the exploration of professions (Q12), the proportion climbs to 30.6%, nearly three times higher, of which 27.8% as a complement and 2.8% exclusively.

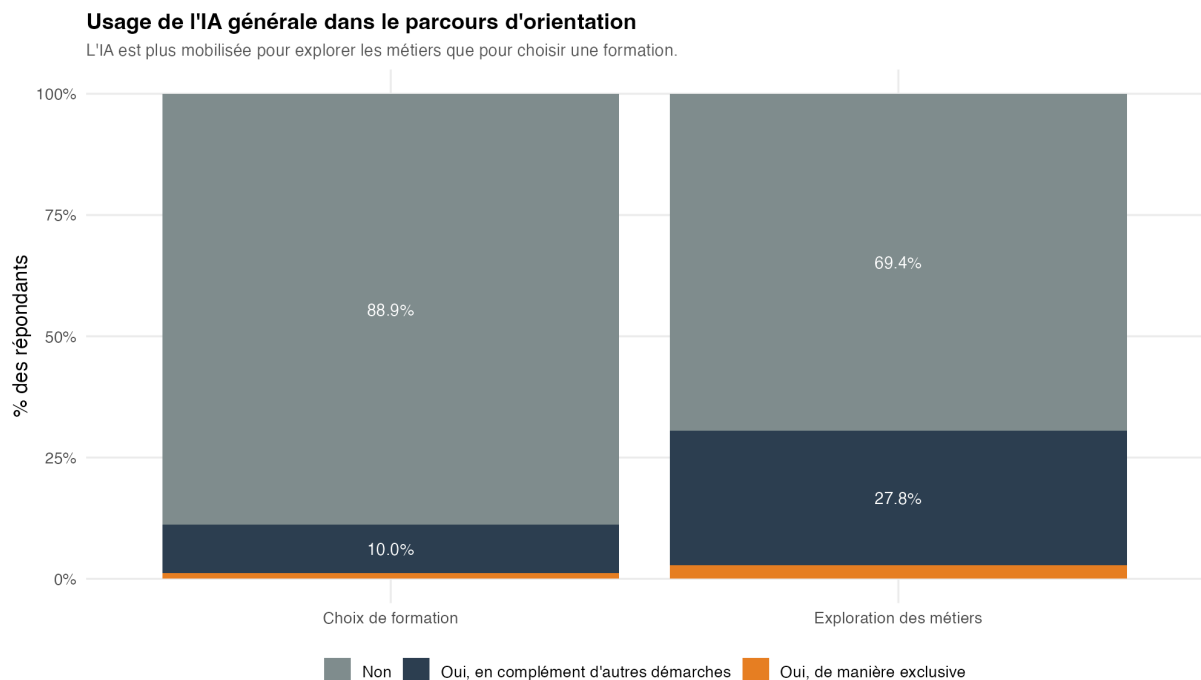


Figure 6 – Use of generative AI in the guidance trajectory. Responses to questions Q11 (use for the choice of programme) and Q12 (use for the exploration of professions). OrientAI pilot, N = 180.

This gap calls for a contextual perspective. Our sample is composed of 96% students at L2 or beyond, that is, of respondents who likely made their post-baccalaureate choices of programme before or at the very start of the mainstream diffusion of generative AI, whose rise dates back to late 2022 with ChatGPT. For some of them, the moment of the choice of programme may not have coincided with the availability of these tools, which could contribute to explaining the low declared usage rate on this dimension, although this cannot be asserted with certainty on the basis of the available data. Conversely, the exploration of professions is a process that continues throughout the student trajectory, including after the emergence of mainstream AI. The differential observed between 11% and 31% could thus reflect, at least in part, a temporal asymmetry between the moment of the initial choices of programme and the present period, rather than a deliberate choice to mobilise AI for one usage rather than another. Our data do not enable a definitive decision between these interpretations.

On the question of a possible social bias in access to these tools, the Chi-square test between parental education level and use of AI for programme choice does not bring out any significant difference ($p = 0.969$). While our data do not enable us to go further on this point, they suggest at least that access to generative AI does not constitute, at this stage, a new line of social inequality in guidance, in contrast to the mechanisms documented in section 3.4.

3.5.2. The use of generative AI does not improve the sense of being informed

The central question for OrientAI is whether the respondents who mobilise generative AI in their guidance trajectory feel better informed than those who do not. The statistical tests deliver a clear negative answer. Concerning the use of AI for the choice of programme, the Kruskal-Wallis test comparing perceived information (Q13) across the three modalities of usage brings out no significant difference ($p = 0.428$). The result is even clearer for the use of AI in the exploration of professions: the mean perceived information is nearly identical between users and non-users ($\mu = 3.18$ against $\mu = 3.12$; $p = 0.984$). The binary versions of these tests, simply comparing users and non-users, confirm the absence of a link ($p = 0.200$ and $p = 0.866$ respectively).

In other words, respondents who have mobilised ChatGPT, Gemini, or another similar tool to inform themselves about programmes or professions do not declare a better sense of being informed than those who have not.

3.5.3. A reading of the observed non-effect

The absence of a link between use of generative AI and sense of being informed calls for a cautious interpretation. One plausible hypothesis is that the generalist tools mobilised by respondents (ChatGPT, Gemini) present their own limitations that may attenuate their capacity to reliably inform on guidance questions. Three lines of explanation can thus be put forward without being able to be settled by our data: the absence of anchoring of these models in French institutional sources (Parcoursup, ONISEP, ROME 4.0), a commercial bias documented in the literature that over-represents private programmes in their recommendations, and the absence of any factual verification mechanism that may lead to inaccurate or outdated information. These limitations of generalist AI, and the way in which the architecture of the OrientAI service seeks to respond to them, are developed in the following sections.

3.6. Synthesis and transversal reading of the pilot

3.6.1. Three central results

Three main results emerge from the pilot.

First, the public guidance service in upper-secondary school is marginal in pupils' actual trajectories. Only 2.2% of respondents cite the guidance counsellor as the principal source of motivation, and 55% never met a PsyEN during their secondary studies. Family and personal circle concentrate 53.6% of the responses. Current guidance thus rests massively on the private informational resources of pupils' immediate circle.

Second, social inequalities manifest themselves through three convergent mechanisms: children from less qualified families tend to report counselling less adapted to their profile (non-significant tendency on our sample but consistent with the literature), orient themselves more towards generalist university tracks ($p = 0.033$), and these tracks concentrate the highest reorientation rates ($p = 0.006$). Although the size of the sample does not enable us to formally test a causal chain, the cluster of converging results is sufficiently clear to underline that inequality of access to quality counselling appears to translate into inequality of placement in higher education, which itself appears to translate into inequality in the stability of the student trajectory.

Third, the perceived quality of the counselling received in upper-secondary school has no directly detectable effect on the probability of reorientation, yet 30% of reoriented respondents consider that better advice could have prevented their change of pathway. Concurrently, the generative AI tools mobilised by approximately one third of respondents to explore professions do not appear to improve their sense of being informed, which invites questioning the fit between existing tools and the need for counselling anchored in French institutional sources.

3.6.2. The moment of reorientation: a peak in L3 to be interpreted with caution

Among the 53 respondents having declared a reorientation, the moment of the change presents a particular distribution. The peak is located in L3 (13 respondents out of 37 whose level could be identified, that is, 35%), followed by L1 (9 reorientations) and then M1 (7 reorientations).

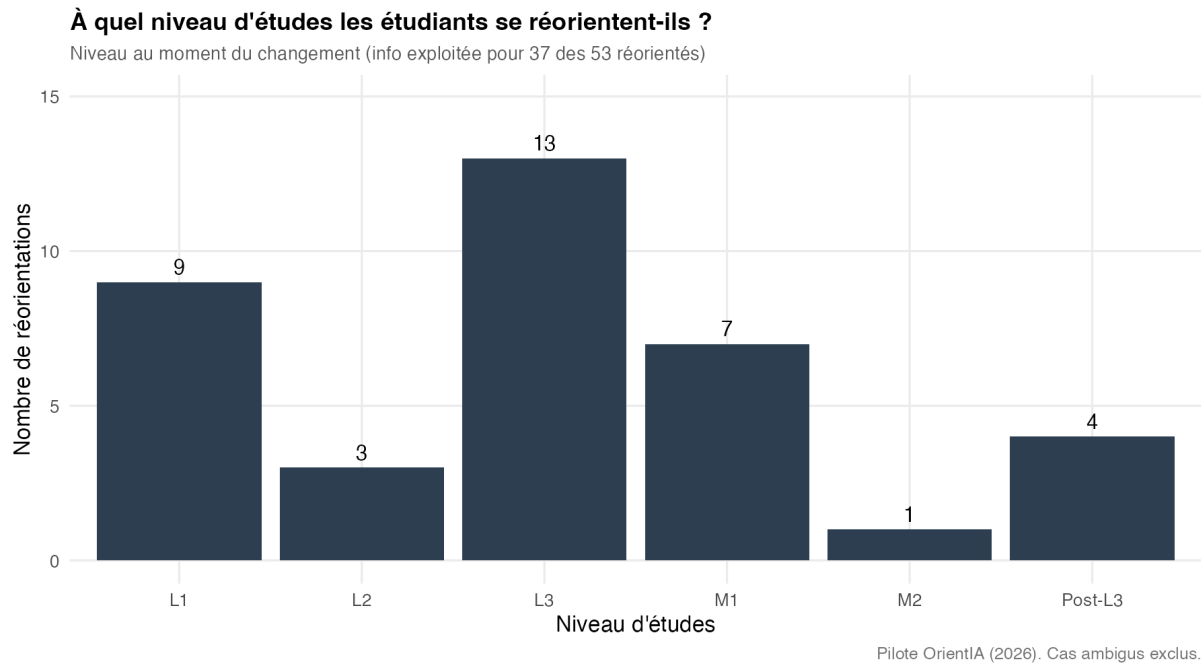


Figure 5 – Level of study at the time of reorientation. Distribution among respondents having declared a reorientation and whose level at the moment of the change could be identified from open-ended responses. OrientAI pilot, N = 37 of 53 reoriented respondents.

This distribution deserves two levels of reading. On the one hand, it contrasts with the French literature, which documents a significant concentration of reorientations at the end of L1. The concentration in L3 in our sample may reflect the favoured academic profile of the respondents, who pass the L1 hurdle without difficulty before realising later that their programme does not correspond to their professional aspirations.

On the other hand, this result raises a definitional limitation that should be made explicit. Our questionnaire does not distinguish changes of disciplinary domain (for example, from economics to health) from changes of institution within the same track (for example, from a university bachelor's degree in economics to a business school). This ambiguity is particularly salient for the peak in L3, which may reflect both genuine changes of domain and strategic transitions between university programmes and private schools. However, the central findings of section 3.4 do not depend on this ambiguity: the 62% reorientation rate observed in Humanities / Social Sciences cannot be explained by prestige transitions, as these tracks have no direct equivalent in schools, and does appear to reflect a change of disciplinary domain in the classical sense.

3.6.3. Overall limitations of the pilot

Beyond the limitations specific to each sub-section, two overall limitations must be explicitly assumed.

The first is the size of the sample (N = 180), which limits the statistical power to detect fine effects and prevents the formal testing of mediation hypotheses between the three dimensions of inequality documented in section 3.4.

The second concerns the temporal scope of the analyses on AI. Our sample is composed of 96% students at L2 or beyond, who for the most part made their post-baccalaureate choices of programme before the mainstream diffusion of generative AI. The analyses on the use of AI for the choice of programme therefore probably underestimate current usage among upper-secondary pupils and recent baccalaureate holders, who constitute the primary target audience of the OrientAI service.

4. Our Solution: The OrientAI Platform

This section describes how the platform works through a representative usage session. The platform can be found here: [OrientAI-platform.fr](https://orientai-platform.fr).

4.1. A Sunday evening

It is 10:12 p.m. on a Sunday in March. Léo is in his final year of an STI2D technological baccalaureate at an upper-secondary school in the Centre-Val de Loire region. He has thirteen days to confirm his Parcoursup applications. On his screen: a Parcoursup tab, an ONISEP tab, an INSA homepage, a forum thread where strangers contradict one another, and a ChatGPT window that he has just closed after it recommended a school that does not exist. His mother is asleep, the CIO opens in two days, and everyone keeps telling him that this is one of the most important choices of his life.

Léo opens OrientAI and types: "I'm in a technological baccalaureate, probably with a 'mention bien', I quite like computer science but I'm afraid the *prépa* will be too hard. What's realistic for me? I'd like to stay in the Centre region if possible."

What follows describes the actual unfolding of his session. By 10:15 p.m., Léo has a substantiated answer. By 10:25 p.m., an appointment with a counsellor is confirmed for Tuesday by video call. Before he closes his laptop, his dashboard will have alerted him to the existence of a grant for which he is probably eligible, and whose very name he was unaware of.

4.2. A three-module architecture

OrientAI is made up of three modules, all accessible from a single user account:

- an AI guidance assistant that answers framing questions by drawing on public sources;
- a network of state-recognised counsellors, to whom the AI defers when a situation falls outside its scope;
- a dashboard that coordinates the administrative deadlines scattered across the various public portals of the student trajectory.

4.2.1. The AI guidance assistant

Léo's first interaction with OrientAI is not a list of programmes. It is a reformulation:

OrientAI: You describe a scientific profile with an inclination towards concrete application, and a legitimate apprehension regarding prépa MPSI. Many holders of a technological baccalaureate turn to other pathways for demanding computer science studies. Three options in the Centre-Val de Loire region, to weigh against what suits you:

The tone remains constant, factual, with no pressure to decide.

No prescription. The AI lays out options, weighs their trade-offs, and flags contraindications. For Léo, it lays out three options: *a BUT in Computer Science at an IUT, a Computer Science Bachelor followed by a MIAGE Master (five years in total), and Polytech via competitive examination after integrated preparation*, each with its target profile, its admission rate, its six-month employment figures, and a direct link to the official programme description. Ranking remains the user's prerogative.

No figure without a source. Every admission rate, every median salary, every employment percentage is tied to a public, dated reference: Parcoursup, ONISEP, InserSup, ROME 4.0, ANSSI SecNumEdu for cybersecurity pathways, RNCP records for professional certifications. The RAG architecture, detailed in section 5, is designed so that the absence of a relevant source in the knowledge base results in a refusal to answer rather than a fabrication.

No prognosis based on socio-demographic profile. A user's origin, gender, or social background is never mobilised to estimate their chances of admission, to restrict the options presented, or to suggest what would be "within their reach". This constraint is built into the system prompt and verified by a set of adversarial evaluations described in section 5.

Léo now knows three things he did not know fifteen minutes earlier: the BUT in Computer Science at Bourges accepts a significant share of holders of a technological baccalaureate, six-month employment rates there exceed 85%, and work-study programmes are available from the second year onwards. The AI ends its response with a question he had not asked himself: *what frightens you, failure or difficulty?*

He does not answer. He books an appointment.

4.2.2. The human relay

Three situations trigger a referral to a human counsellor.

Trade-offs that are not factual in nature. Some guidance decisions cannot be resolved by better access to information. Choosing between an excellent programme far from home and an adequate one close to one's family engages a hierarchy of values that is specific to each individual: academic ambition, emotional proximity, residential autonomy, none of which is settled by any employment figure. Likewise, committing €9,000 per year to a private school is not a question of abstract return on investment but a decision that bears upon an entire household's budget, and that calls for conversations that only the user can have. OrientAI makes available the elements that illuminate these decisions, without claiming to make them on anyone's behalf.

Explicit request from the user. A "Book an appointment with a counsellor" button is accessible at any point in the conversation, without any requirement to justify the request.

markers of distress detected in conversation. Social isolation, expressions of distress, doubt about the very meaning of pursuing studies: these markers trigger, in addition to a referral to a counsellor, the display of the national helpline numbers (3114, Fil Santé Jeunes). The detection of these signals is subject to dedicated evaluations in section 5.

The transition from the AI to the counsellor is accompanied by a synthetic briefing transmitted ahead of the appointment. Its generation follows three rules.

Explicit consent. No element is transmitted without the user's validation on a dedicated screen, which displays the full content of the briefing before sending.

Synthesis, not transcript. The briefing is not the raw history of the conversation, but a structured summary: declared academic profile, options under consideration, constraints that the user wishes to share, and open questions that the conversation did not resolve.

Free editing. Any element may be removed or rephrased before transmission.

When Léo connects on Tuesday at 2:00 p.m., his counsellor has read his trajectory. The appointment begins at the first minute, not the fifteenth.

OrientAI does not substitute itself for counsellors. It concentrates their time on the conversations that require it, and ensures that those conversations begin informed.

4.2.3. Temporal continuity

The administrative pathway of a French student is not a multi-layered set of forms to be filled in parallel. It is a nine-month calendar, spread across several public portals, which no single source consolidates.

The problem is not that these steps are numerous; it is that they are spread out. In July, a student who began submitting applications in January must remember a file left pending in March. What fails is continuity.

The dashboard replaces no portal; it coordinates them. A single page displays what must be done this week, what is to be prepared for next month, and what will arrive later in the year. Notifications are calibrated according to the criticality of each deadline: a reminder at D-14 for the DSE, a daily reminder at D-3 for the Parcoursup confirmation, a single alert for the CVEC.

For Léo, on this Sunday evening, the dashboard foregrounds the imminent confirmation of his applications. Above all, it surfaces a deadline he had not seen coming: the Dossier Social Étudiant, to be submitted before 31 May. Given the family situation he described to the AI, he is likely eligible for a high-level grant and for priority CROUS housing. Each year, tens of thousands of students miss this window and start their first year without an entitlement they were eligible for.

Léo had not forgotten the DSE. He did not know it existed.

4.3. Accessibility

The experience described above must work for all young people in France, and not only for those who match Léo's profile. The platform incorporates, by design:

RGAA 4.1 compliance (based on WCAG 2.1 AA): the French reference framework for digital accessibility, mandatory for public services and adopted voluntarily here.

Full keyboard navigation: every action can be performed without a mouse or a touchscreen (skip-link, Tab navigation, panel closure by Escape). Essential for users with motor disabilities or those using a screen reader.

Structured announcements for screen readers: the conversation area uses the ARIA live regions (role="log", aria-atomic) supported by NVDA, VoiceOver, and TalkBack. User tests with these tools are planned.

Dyslexia mode: OpenDyslexic font and spacing adapted to specific learning differences (British Dyslexia Association recommendations), activatable in one click. High baseline contrasts (AAA text, AAA primary).

Simplified reading mode: enlarged typography, generous spacing, clickable areas of 48 pixels. Automatic FALC reformulation of responses currently under development.

Audio playback and voice dictation: responses can be listened to and questions asked aloud rather than typed, in French.

For Léo's session to unfold as described, and for it to unfold identically for any other user, these three modules must honour several commitments: systematic sourcing of responses, refusal rather than fabrication in the event of a documentary gap, controlled transmission of context to counsellors, and compliance with the requirements of Annex III of Regulation (EU) 2024/1689. These commitments rest on precise technical choices: a sovereign model, a RAG knowledge base built on French public sources, a validation pipeline, and documented human oversight. Section 5 details their design.

5. OrientAI: architecture of the conversational module and benchmark

This section describes the technical design of OrientAI's conversational module: the choices that distinguish it from the generalist LLMs currently used by half of French students, the architecture that makes it factually grounded, the safeguards that frame its responses, and the limitations we explicitly assume before the transition to the concrete conditions of deployment developed in section 6.

Before entering into the technical detail, a framing precision is in order. What follows was designed and developed over one month, by a two-student team, on personal machines (8 GB of RAM). This material constraint is not an incidental detail: it has structured each of our architectural choices, and it is also what we wanted to put to the test. Our intention was to verify that a useful, sourced, and equitable tool could be built at this scale, with these means, and we leave it to the reader to judge the outcome.

5.1. Why a system grounded in French sources rather than a generalist LLM

The literature review (section 2) documented a series of converging weaknesses of the generalist LLMs currently mobilised by young people for guidance purposes. Their training on the global web leads them to overrepresent the marketing content of private programmes at the expense of the public landscape structured by Parcoursup, ONISEP, and RNCP records. They moreover reproduce social and gender biases documented by recent audits, going so far as to adjust the difficulty of the content produced according to the user's presumed family income, independently of their actual abilities. Added to these biases are two architectural limitations: the absence of any factual verification mechanism, which makes a fabricated admission rate and a real one indistinguishable in their output; and a knowledge frozen at the training date, which poses a particular problem in a domain where Parcoursup, MonMaster, and RNCP records are updated annually.

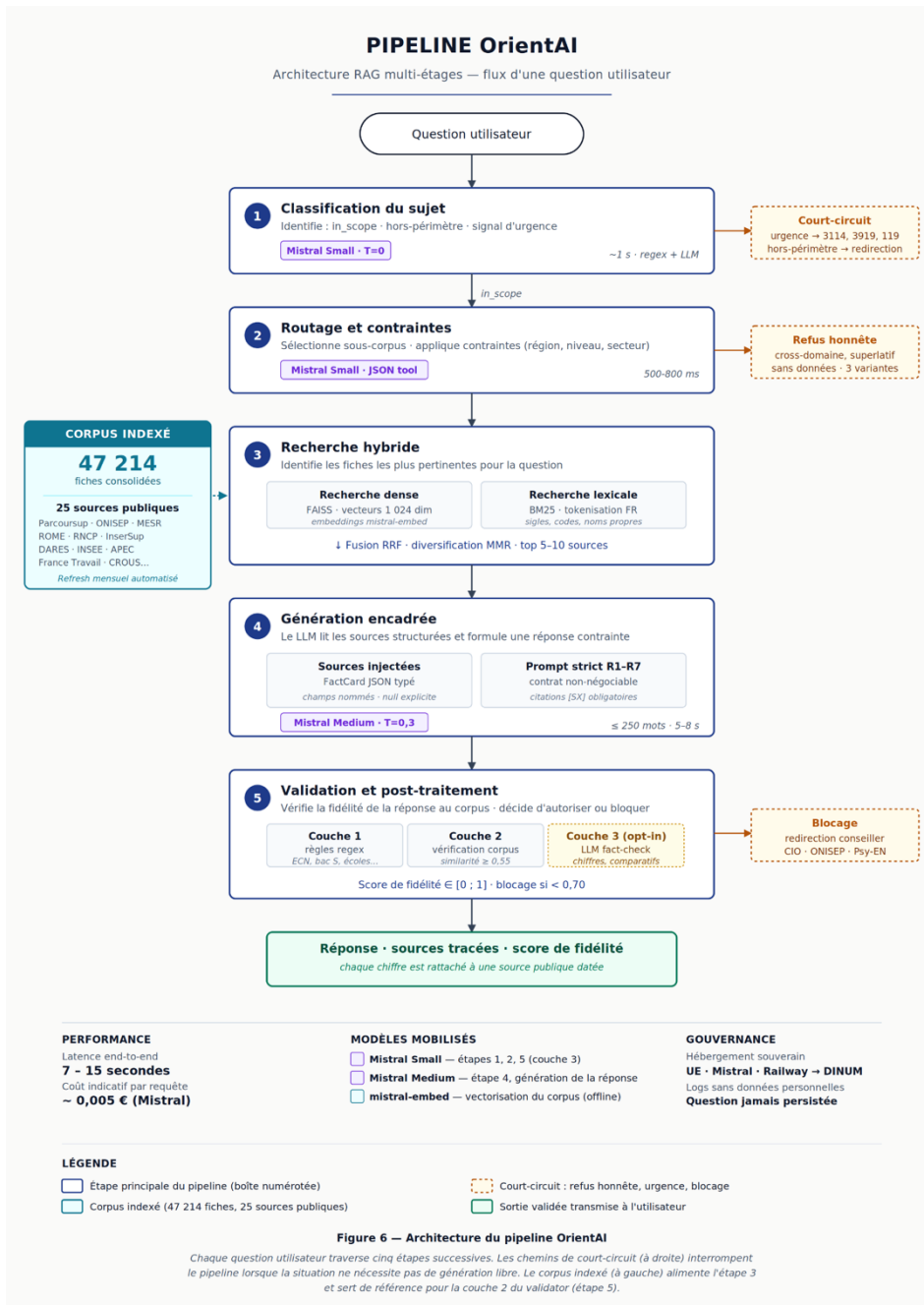
These weaknesses call for a specific architecture: a system that separates knowledge (a corpus of official French sources, controlled and maintained) from generative capacity (a language model that formulates the response). This is what the literature calls a RAG architecture (Retrieval-Augmented Generation), now standard in LLM deployments on specialised domains where accuracy takes precedence over creativity (Lewis et al., 2020; Gao et al., 2024).

The choice of a RAG architecture rather than fine-tuning the model on guidance data follows from a constraint specific to our domain: Parcoursup descriptions, InserSup employment rates, and MonMaster calendars change every year. Retraining a model on each update would be costly and slow; refreshing an indexed corpus is trivial. This property will be decisive for the move to national scale.

The choice of Mistral as the provider of the language model responds to three cumulative requirements: the sovereignty of the data processed by the model, native French quality (public benchmarks establish that Mistral models match or exceed the performance of equivalent American models on French-language tasks), and compliance with European requirements. Mistral is a French provider hosted in Europe, which simplifies the compliance analysis with the AI Act and the GDPR.

5.2. Pipeline overview

A user query passes through several successive stages, designed to minimise the risk of hallucination at each link. The figure below summarises the flow of a question from the interface through to the response.



Architecture of the OrientAI pipeline. Five stages frame the generation produced by the Mistral Medium model. Each stage is designed to reduce the risk of hallucination or to short-circuit the LLM when the question pertains to a situation that does not require free generation (emergency, precise factual lookup, out-of-scope).

The counter-intuitive aspect of this architecture deserves to be emphasised: the objective is not for the LLM to answer every question, but for it to refuse honestly when the corpus does not contain the relevant information. Stages 1, 2, and 5 all contain short-circuit paths that interrupt the pipeline before generation if the situation does not lend itself to it. This is the structural difference with a generalist LLM, whose objective function is to produce a plausible response regardless of the state of its internal knowledge.

5.3. The corpus: 47,214 records from 25 official public sources

The quality of a RAG system is conditioned by that of its data. Our corpus, consolidated as of 9 May 2026, aggregates 47,214 records from 25 official public sources, normalised, deduplicated (merging of identical records across sources), and enriched. The table below presents the main sources mobilised.



Three properties of this corpus deserve comment. First, institutional diversity: no single source accounts for more than 18% of the corpus, which reduces the risk that a bias specific to one source might contaminate the whole system. Second, relevance: the corpus is refreshed monthly by an automated ingestion pipeline ([src/collect/](#)) that tracks updates of the original sources; this property, impossible to obtain with a generalist LLM whose knowledge is frozen, gives the system a structural advantage on questions that depend on recent data (admission rates from the latest campaign, ongoing calendars). Finally, each record retains the trace of its source, its extraction date, and its canonical URL, which enables post-hoc auditing at any time.

5.4. Searching within the corpus: vector indexing and hybrid retrieval

When a question arrives, the system must identify within a few hundred milliseconds the most relevant records among the 47,214 available. This operation mobilises two complementary retrieval mechanisms, whose combination has become standard in production RAG systems (Cormack et al., 2009; Mandikal & Mooney, 2024).

Dense retrieval relies on the vector representation of records. Each record is translated by the mistral-embed model into a 1,024-dimensional vector that captures its semantic meaning. The user question is translated in the same way, and the system identifies the records whose vector is closest to that of the question. This retrieval, indexed by the FAISS library developed by Meta AI, makes it possible to find semantically related records even when the exact words differ (a question such as "I want to work in human rights advocacy" returns records for *avocat*, *juriste*, *magistrat* although none of these terms appears in the question).

Lexical retrieval (BM25) operates in complement on exact words. It is essential for queries that contain acronyms, proper nouns, or codes (an "INSA Lyon", a "code ROME M1805", a "BUT GEII") that dense retrieval may dilute into approximate semantic neighbourhoods. The fusion of the two rankings by *Reciprocal Rank Fusion* (Cormack et al., 2009) combines the strengths of each approach: a document appears in the final top-K only if it is ranked highly by at least one of the two methods.

The corpus is furthermore partitioned into four *thematic sub-indices* (programmes, professions, statistics, regional support schemes) which are selectively activated according to the subject of the question, as identified by an upstream router. This partition reduces cross-domain pollution: a question about the remuneration of a profession does not accidentally bring up programme records at the top of the ranking.

Lastly, the *MMR diversification* algorithm (Carbonell & Goldstein, 1998) is applied to the final candidates to prevent the top-K from being composed of five very similar variants of the same programme, at the expense of relevant alternatives. This detail matters in the guidance context: an upper-secondary student asking for "BUTs in computer science in Île-de-France" deserves to see five different institutions, not five records from the same one.

In summary, the retrieval pipeline identifies in under a second the dozen or so records most relevant to a given question, by combining the semantic richness of dense retrieval with the lexical precision of BM25.

5.5. Framed generation: the strict contract imposed on the LLM

Once the sources have been selected, the generator LLM (Mistral Medium) receives two distinct blocks. The first is a strict system prompt that defines its role, its obligations, and its prohibitions. The second is the set of selected sources, formatted not in free prose but in typed JSON (which we call FactCard) where each field is named and each value made explicit.

```

{
  "id": "S1",
  "formation": "BUT Informatique",
  "etablissement": "IUT Lyon 1",
  "ville": "Villeurbanne",
  "region": "Auvergne-Rhône-Alpes",
  "niveau": "bac+3",
  "statut": "public",
  "type_diplome": "BUT",
  "chiffres": {
    "taux_acces_parcoursup_2025": 12.4,
    "nombre_places": 60,
    "duree": 3,
    "frais_annuels": 170,
    "taux_emploi_3ans": 0.92,
    "salaire_median_embauche": null
  },
  "url": "https://dossierappel.parcoursup.fr/ ... ",
  "annee_donnees": 2025,
  "provenance": {
    "tier": "tier_1",
    "source_label": "Parcoursup",
    "last_updated": "2026-04-30"
  }
}

```

This choice of injecting sources in JSON rather than in free prose is central. When an LLM reads prose, it can paraphrase, summarise, and, this is the risk, interpolate information that did not appear there. When it reads typed JSON, each field is explicitly separated: a *taux_emploi_3ans* field with a value of null sends an unambiguous signal that this data is not available for this source. The system prompt exploits this structure by stating a non-negotiable rule: « You may ONLY cite values present in the *chiffres* block of a source. Any other numerical value is FORBIDDEN. If the field is null → 'information not available in my sources'. »

The system prompt contains seven numbered rules (R1 to R7) covering figures, the identity of programmes, source citations, style, posture, length (the longer the response, the higher the probability that the model hallucinates), and the contextual constraints injected by the router. Four of the seven rules are reproduced below for illustration:

R1: *Figures*: You may ONLY cite the values present in the *chiffres* block. If the field is null → "information not available in my sources".

R2: *Identity of programmes*: You may ONLY cite programmes whose triplet (programme + institution + city) appears in <sources>. Empty sources → "I have no relevant programme in my sources; I recommend you book an appointment at the CIO".

R3: *Citations*: Every cited figure MUST be followed by a reference of the form [source S1].

R5: *Posture*: Empathetic without overplaying it; direct if the project is not realistic; no discrimination; final open question.

This explicit-contract approach is documented in the recent literature as one of the most effective techniques for reducing the hallucination rate of LLMs on factual tasks (Wei et al., 2022; Tonmoy et al., 2024). It does not eliminate the risk but substantially reduces it, especially when combined with the downstream validation mechanisms described below.

In summary, the LLM never generates from its own memory: it reformulates, within a constrained stylistic and length envelope, the content of a typed JSON that has been supplied to it, without being able to depart from it.

5.6. Anti-hallucination safeguards: a multi-layered apparatus

Even with a strict prompt and sources injected in typed JSON, an LLM can produce a response that appears credible but departs from the corpus. The risk is asymmetric: in a guidance domain, a single fabricated figure or a single phantom programme can durably steer a young person's decision. We have therefore placed an automated safety net after each generation, whose function is not to produce a good answer, this is the role of the preceding stages, but to detect and block those that drift from the corpus.

This net rests on two complementary layers, executed within a few dozen milliseconds before the response reaches the user.

The first layer tracks known factual confusions. It consists of a dozen hand-written rules, each targeting a recurring error observed in Mistral Medium's outputs: mention of the "ECN" (renamed EDN since the 2023 reform), of the "bac S" (abolished by the Blanquer reform), of an "École 42 free under work-study" (factually incompatible with the school's pedagogical model), or of a pathway from "humanities bachelor's degree to speech therapy" (non-existent in the French system). These rules do not cover every possible case, but they catch the most frequent errors.

The second layer verifies that the cited programmes actually exist. For each programme mentioned in the response, the system searches for a match within the indexed corpus, taking into account both the name of the programme and the institution that delivers it. If no record in the corpus resembles what the response cites closely enough, the programme is flagged as suspected of being fabricated. In practice, this layer is the most discriminating: it detects entirely fabricated programmes, which constitute the most damaging class of error because it is the hardest for a non-expert user to detect (a plausible but fictitious bachelor's degree name passes for true in the eyes of an upper-secondary student discovering the landscape).

The aggregation of these two layers produces a fidelity score between 0 and 1, which drives a three-tier response policy. Above 0.70, the response is deemed reliable and transmitted to the user. Between 0.55 and 0.70, it is transmitted but accompanied by an explicit warning regarding the elements to be verified. Below this threshold, the response is replaced in its entirety by a message inviting the user to book an appointment with a human counsellor: this is the ultimate safety net, which favours honest refusal over taking risks.

A final stage, purely deterministic, cleans up surface artefacts: it removes hyperlinks to non-existent sites that the LLM may occasionally fabricate, corrects malformed tables, and verifies that the cited ONISEP record identifiers actually exist within the mobilised sources. These operations never modify the factual content of the response; they only eliminate the formal flaws that could mislead the user.

In summary, two independent layers verify that the response remains faithful to the corpus, and three thresholds make it possible to transmit, warn, or block according to the level of confidence.

5.7. Initial benchmark: comparison with generalist LLMs

To evaluate whether the architecture described so far produces responses that are genuinely better than those of a generalist LLM queried without a domain corpus, we built a multi-layered evaluation protocol. Seventy-one questions were retained, distributed across eleven categories (upper-secondary students in their final year, students considering reorientation, profession discovery, geographical constraints, student life, adversarial queries, medical out-of-scope queries, paraphrases without institutional vocabulary). Each question was posed in parallel to seven systems: OrientAI v4.1 on one side, six configurations of generalist LLMs on the other, combining Mistral, GPT-4o and Claude in neutral mode or with a structured prompt without corpus. The responses were then anonymised, permuted and submitted to two external judges (Claude Sonnet 4.5 and GPT-4o), each rating every response on a rubric of six criteria: institutional neutrality, realism of figures, sourcing, geographical diversity, agency, and discovery of less well-known pathways. The main result concerns the two axes that this proposal explicitly defends. On sourcing, OrientAI v4.1 achieves a score markedly higher than any baseline tested. It is the widest gap measured across the six criteria, and it is expected: the mandatory citation rule after every figure, combined with the corpus verification layer, is by construction what no generalist LLM produces. On institutional neutrality, that is, the valuing of public labels and the absence of any premium granted to commercially more visible private schools, the gap is more modest but remains in OrientAI's favour against each of the neutral baselines. These two results empirically validate the two most structuring architectural choices of the proposal: the injection of sources in typed JSON with mandatory citations, and the prioritisation of state-recognised institutional labels in the reranker.

The second strong result concerns the calibrated refusal of out-of-scope questions. Across the thirteen adversarial and out-of-scope requests in the protocol (fake schools, fictitious dates, prompt injections, purely medical queries), OrientAI v4.1 refuses correctly in twelve cases out of thirteen, that is 92.3%. The rate reaches 100% on the purely out-of-scope questions. This property, which none of the generalist LLMs tested matches under comparable conditions, is the empirical translation of the apparatus described in 5.6: the ScopeClassifier and the RouterLLM jointly produce an asymmetry in the face of uncertainty which, from an upper-secondary student's standpoint, is exactly what one expects from a sourced tool, namely to refuse rather than fabricate.

The cost of these choices, measured by the same protocol, must now be set out. The production constraints of the v4.1 strict prompt, namely responses capped at 250 words and regional locking on questions carrying an explicit geographical constraint, penalise two criteria of the rubric: geographical diversity and discovery of less well-known pathways. On these two dimensions, generalist LLMs in neutral mode obtain higher scores than OrientAI v4.1, and the gap is appreciable. This result is not a defect, in the sense that it signals no error; it measures the cost of a deliberate product choice. A talkative LLM that proposes twenty-five pathways scattered with unverifiable figures obtains a better mark on "narrative breadth" than a system that responds strictly within the requested region with three sourced options. What constitutes a quality from the rubric's standpoint is precisely what our architecture seeks to avoid from the

user's standpoint. We document the gap rather than minimise it, because it falls to the reader to judge whether the trade-off is defensible.

A less comfortable result deserves to be set out directly. The two external judges diverge substantially in their overall assessment. Claude Sonnet validates the proposal's thesis by placing OrientAI v4.1 markedly above the neutral baselines; GPT-4o invalidates it by awarding Mistral neutral a score markedly higher than every other system tested, including those issued by its own vendor. This asymmetry is not a coincidence: GPT-4o tends to reward exhaustive coverage and free-flowing prose, whereas Claude rewards explicit sourcing and concision more heavily. The disagreement is itself a result. It indicates that the quality of a guidance response is not a univocal magnitude, and that evaluating a system such as OrientAI presupposes an explicit choice of what one calls "good". We document this disagreement rather than retaining only the favourable judge, because opacity on this point would be precisely the kind of bias this proposal reproaches in existing systems.

In summary, OrientAI v4.1 outperforms generalist LLMs on the two criteria this proposal defends, namely verifiability of figures and institutional neutrality, and correctly refuses out-of-scope questions in nine cases out of ten. This advantage carries a measurable cost on criteria of narrative breadth that evaluation rubrics traditionally reward, a cost we assume as the direct consequence of a product choice. The disagreement between our two external judges on the overall assessment is set out as such: it serves as a reminder that the quality of a guidance tool is not measurable independently of the criterion one chooses to privilege.

5.8. Data governance and regulatory compliance

Guidance for young people is one of the use cases that the European AI Act, in progressive force since 2024, explicitly classifies among high-risk AI systems. This classification, to which Christelle Scharff (Pace University, scientific committee of the competition) drew our attention, triggers reinforced obligations regarding the quality of training data, the transparency of operation, the traceability of decisions, and effective human oversight. Rather than treating these obligations as a compliance constraint added after the fact, we have integrated them from the design phase as structuring principles of the architecture.

Sovereign hosting is ensured by the choice of Mistral as the model provider. User queries and responses transit through a European infrastructure, which eliminates the principal flaw of the American LLMs used by default today: the transfer of particularly sensitive data (academic results, professional aspirations, family circumstances of a minor or young adult) to foreign operators outside any national educational framework. On the service infrastructure side, the deployment is carried out on Railway for the prototype phase, with an identified trajectory towards OVHcloud Confidential or DINUM operators for the production phase, both of which are SecNumCloud-qualified.

The processing of user data follows a strict minimisation principle: the user's question is never logged in the server logs; only an anonymised request identifier, latency, the number of sources mobilised, the fidelity score, and the length of the question (without its content) are retained. This technical decision, explicitly encoded in the production code ([src/api/server.py](#)), guarantees that a leak of the logs would reveal no personal data and makes it possible, by construction, to respond to the GDPR rights of access, rectification, and erasure (Article 15 et seq. of the GDPR).

The audit of sources prior to indexing is the aspect on which Christelle Scharff's reminder was most structuring for our reasoning. A RAG system, by construction, restitutes what its corpora contain, and if these corpora already reflect structural inequalities in the way programmes and professions are described or ranked, the system can reproduce these biases under an appearance of technical neutrality. This warning led us to inscribe in the project a stage of auditing the mobilised sources before indexing, which is not limited to their regulatory compliance but also examines their representativeness (for example, the documented under-coverage of overseas institutions in national databases) and their potential handling of sensitive categories. This stage, currently carried out by hand on the 25 existing sources, will need to be formalised and documented as a standard deliverable for every new source integrated, and submitted to external oversight; section 6 proposes that this oversight be entrusted to INETOP through the CNAM, already a member of the AVENIR(s) consortium.

The effective human oversight required by the AI Act for high-risk systems is implemented in our architecture by the counsellor relay module (cf. section 4.2.2). At no point does OrientAI take a decision in the place of the user or the counsellor: the system informs, sources every assertion, and proposes a connection to a human counsellor for all trade-offs that exceed the factual scope. The traceability of decisions is ensured by the persistence of the sources mobilised for each response (visible on the API side in the `sources` field of the return), which enables a counsellor, an auditor, or the user themselves to trace from the advice received back to the sources that supported it.

5.9. Acknowledged limitations and conditions for national scaling

This section describes what the current prototype system does not do, or does only partially, and what would be required for it to become a public AI guidance service that could effectively be deployed at national scale. Methodological honesty requires distinguishing scope limitations, which can be lifted by extension of the corpus, from architectural limitations, which presuppose a deeper redesign, and from organisational limitations, which can only be resolved within an appropriate institutional framework.

Regarding the scope of the corpus, several areas remain insufficiently covered. The agricultural track (Parcoursagri) is not yet integrated, which deprives the system of approximately 1,200 specific programmes. Private programmes outside the RNCP, which represent a significant share of the post-baccalaureate landscape, are not integrated by default so as not to overvalue pathways whose institutional quality is not guaranteed; the inclusion criterion (public label, presence in an official database) will nevertheless have to be formally documented. Overseas programmes are underrepresented (16 DOM-TOM records in the current corpus), which reflects a gap in the original national databases but which the project cannot simply pass on. Adult reconversion schemes (CPF, AFEST, VAE) are only partially mapped, which limits the system's usefulness for the adult audience for the time being.

On the architectural level, the current system is single-turn: it answers one question at a time, without building a persistent model of the user across conversations. This property is deliberate at the prototype stage because it simplifies GDPR compliance analysis and limits attack surfaces, but it reduces the usefulness of the service for long-term support, which presupposes at the very least a memory of the user's trajectory: previous application choices, declared constraints, administrative deadlines. The dashboard module described in section 4.2.3 provides for this persistence; its coexistence with the conversational module remains to be designed and audited from a data protection standpoint. Likewise, retrieval does not use contextualised query

rewriting (HyDE techniques, query expansion): the current system searches from the user's words, which may limit the quality of retrieval for indirectly phrased questions or for users who are not familiar with institutional vocabulary.

On the evaluation side, the protocol described in 5.7 rests on seventy-one questions assessed by two external LLM judges, which constitutes a proof of methodological feasibility but remains insufficient for any statistically conclusive claim. Three extensions are necessary before any deployment: an expansion of the evaluation set to several hundred questions, covering more finely the underrepresented categories (adult reconversion, overseas territories, vocational tracks); a structured user test conducted with upper-secondary students and guidance counsellors, in order to measure the system's actual usefulness rather than its indirect score; and a longitudinal evaluation procedure that follows a single user across the duration of a guidance campaign rather than on an isolated question.

On the organisational level, finally, several structuring mechanisms remain to be built. A public mechanism for reporting biases and errors, modelled on pharmacovigilance, would enable any user or counsellor to flag a problematic response and would trigger a documented audit procedure. A formalised procedure of biannual auditing of the system's sources and biases, conducted by an independent institutional actor (INETOP being the natural candidate through its attachment to the CNAM), would guarantee long-term oversight that a design team cannot provide. Training counsellors in the use of the tool, articulated with the IH2EF, conditions the quality of the human relay that constitutes the ultimate safeguard of the system: a counsellor poorly informed about the tool's limitations might, paradoxically, grant it excessive confidence. Lastly, integration with existing portals (Parcoursup, MonMaster, AVENIR(s)) presupposes technical and institutional agreements that go beyond the scope of a prototype and constitute the proper subject of the following section.

These limitations do not disqualify the proposed architecture; they delimit its current prototype status and indicate the work that the deployment phase will have to open up. The project's wager is precisely that these tasks only arise because a technical proof of concept exists, and that they can only be resolved within the framework of an appropriate institutional anchoring: this is the subject of section 6.

6. Implementation: institutional anchoring, deployment, governance

The preceding sections have set out the diagnosis, the proposition, and the technical architecture of OrientAI. It remains to examine the concrete conditions under which such a project could be taken up by public authorities. We are aware that the precise definition of this arrangement falls less to a choice of students than to an institutional dialogue that it will be for the public authorities to initiate. Our intention is more modest: to share the observation of a gap and to sketch a student response to a student problem. We move daily within the university environment, where we observe both the needs of our peers and the actual uses they make of generative AI. It is from this position that we formulate this proposal, in the hope that it may nourish the debate and contribute to making more visible a reality that, to this day, remains largely under-addressed by existing public tools.

Participation in this competition has also provided a working framework that exceeds the production of deliverables. The keynotes, the resources made available by the organisation, and the exchanges with other teams have nourished our reflection on several occasions and led to

recalibrations that we mention throughout this section, wherever they have caused our reasoning to evolve.

6.1. Institutional and economic anchoring within the AVENIR(s) programme

The question of institutional anchoring conditions the legitimacy of the service in the eyes of pupils and families, its durability beyond a one-off initiative, and its articulation with state-recognised guidance counsellors. No private solution can hold this twofold promise of neutrality and durability. It remains to be seen which public institution would be best placed to take up the project.

Our initial reflection focused on three institutions liable to play this role, each identified for a different form of legitimacy. The ONISEP held the obvious operational legitimacy, as the public reference operator for guidance and as the holder of the databases that our prototype already mobilises (programme descriptions, profession descriptions). The INETOP, a century-old institute of the CNAM dedicated to research in guidance and to the training of National Education psychologists, brought the scientific endorsement required by a project mobilising LLMs in a sensitive domain. The IH2EF, a service of national competence attached to the ministries of National Education and of Higher Education and Research, trained school principals and CIO directors, which ensured the operational anchoring among the executives who in practice decide on the deployment of tools in institutions. At the outset of our reflection, we were moving towards an arrangement with three independent overseeing bodies.

It was in deepening the analysis of the ONISEP as potential bearer of the project that we discovered the existence of the AVENIR(s) programme, whose scale and maturity substantially modified our approach. Coordinated by the ONISEP with a consortium including the CNAM, Université Savoie Mont-Blanc, Université de Lille, the CEREQ, and the CRI, financed to the tune of €30 million over 10 years within the framework of France 2030 (French Government, 2022), AVENIR(s) is a digital platform for education on guidance, deployed in all French *collèges* and *lycées* since September 2025. It offers pupils a skills portfolio, an agenda of events, interactive pedagogical activities, and a personal space to structure their reflection, also accessible to educational teams and to National Education psychologists. The *Plan Avenir* announced by Minister Élisabeth Borne in June 2025 broadened its ambition, with an explicit objective of reducing inequalities in trajectories linked to social origin, gender, or territory (Ministère de l'Éducation nationale, 2025).

The detailed examination of AVENIR(s) brought to light several points of convergence with our proposal, the scale of which we had not initially measured. The diagnosis is largely shared: the *Plan Avenir* observes, as we do, that more than 60% of parents consider information on programmes to be insufficient and that one upper-secondary student in two does not feel well supported. The architecture is in part common: three of the principal functions that our project envisages (structured monitoring of the trajectory over time, centralisation of deadlines, articulation with human support by state-recognised counsellors) are already sketched in AVENIR(s) in different forms. The logic of governance shared between the State, the Regions, public operators, and educational institutions is likewise asserted by the *Plan Avenir* as by our proposal.

A comparative analysis nevertheless brings to light several areas that the current platform does not cover and that OrientAI would precisely come to invest. The first, and most structuring, is

the absence of a conversational generative AI component. AVENIR(s) offers editorial resources and pedagogical activities structured over time, which corresponds to a logic of pedagogy of guidance. But none of these tools enables a young person to freely ask a question, in their own words, and to obtain a personalised and sourced answer. It is precisely this need that young people today address to foreign generalist LLMs, outside any educational framework, as the literature review has documented. A French conversational layer, sovereign and compliant with European requirements, remains to be built within the public guidance architecture.

The second area not covered concerns audiences. AVENIR(s) is designed for school time and addresses *collège* and *lycée* pupils. Yet guidance needs do not stop at the exit from upper-secondary school. Students faced with a reorientation choice, young adults in professional reconversion, and people removed from educational structures who wonder about bridges or short programmes currently have no public digital service designed for them. OrientAI is explicitly conceived to cover this continuity, by addressing anyone asking a guidance question, regardless of their academic status and the moment at which the question arises.

We therefore revised our approach to institutional anchoring towards a composite arrangement articulated around AVENIR(s): operational anchoring by the ONISEP as an extension of the existing platform, which would pool the databases and the infrastructure; scientific endorsement provided by the INETOP through the CNAM, already a member of the AVENIR(s) consortium, to validate the benchmark protocols and supervise the content produced by the AI; and an articulation with the IH2EF for the training of executives and counsellors in the use of the tool. This arrangement reflects the reality of the French institutional landscape, in which diffusion, research, and the training of executives fall to distinct institutions, and it fits within the logic of shared governance asserted by the *Plan Avenir*. The institutional window for such an anchoring is moreover open: the *Plan Avenir* explicitly anticipated an evolution of the operator in 2025, the director general of the ONISEP being mandated to submit to the Government, by mid-July, proposals for the evolution of the operator (Ministère de l'Éducation nationale, 2025). The context is therefore one of active institutional reflection on the evolution of the operator, within which the integration of a generative AI component would have its full place.

This anchoring is also what makes the project's economic model tenable. A public AI guidance service is viable only if it is designed with a progressive deployment trajectory and a sustainable economic model, and these two dimensions are inseparable: it is the progressivity of the deployment that makes the cost absorbable, and it is the mastery of the cost that makes the progressivity credible. The deployment of AVENIR(s) offers a useful precedent. Launched in December 2021 within the framework of France 2030, the programme went through four successive beta-testing phases with pilot institutions, the last of which involved 340 *collèges*, *lycées*, and CIOs distributed across all academies (Café pédagogique, 2024). The progressive opening to upper-secondary students and then to lower-secondary students took place in late 2024, and the generalisation to all institutions did not occur until the 2025 school year (Ministère de l'Éducation nationale, 2025). Nearly four years between launch and national deployment. This progressivity made it possible to adjust the tool to the actual practices of educational teams, to stabilise the infrastructure, and to absorb costs over time. The same logic applies to OrientAI: a controlled deployment across a few pilot academies in liaison with the CIOs, SCUIOs, and partner *rectorats*, then a regional broadening progressively integrating the complementary modules, and only in the last resort a national generalisation, once the prior phases will have validated the quality of the service. No AI system applied to a sensitive domain can be deployed all at once across an entire country without major risk.

The costs associated with this deployment are concentrated on three items: technical infrastructure (sovereign hosting, inference capacity, vector databases), scientific editorial work (updating of data, validation of benchmarks, supervision of content, correction of biases), and articulation with human counsellors. The envelope for infrastructure alone would be on the order of a few hundred thousand euros per year. Several orders of magnitude in the public sphere make it possible to frame the budgetary discussion: since 2024, the DINUM has devoted an annual budget of €1.2 million to all of its generative AI projects, including Albert, the State's AI assistant developed with Mistral and hosted by Outscale (Weka, 2026); the ONISEP holds a subsidy for public service charges of €22.7 million in the 2025 finance bill (Sénat, 2024-2025); and the AVENIR(s) programme benefits from €30 million in funding over 10 years within the France 2030 plan (French Government, 2022). These references situate the budgetary scope within which a service such as OrientAI could reasonably be inscribed.

Rather than building a service *ex nihilo* whose entirely new budget would need to be justified, OrientAI proposes to graft a generative AI component onto a public infrastructure currently being deployed, one that already has its databases, its network of counsellors, its shared State-Regions governance, and a secured multi-year budget. The cost of OrientAI thus comes down to two items: developing the AI component and evaluating it scientifically. This logic of pooling, combined with a tiered deployment that adjusts costs to the maturity of the service, is the condition for a sustainable economic model.

6.2. The human conditions of deployment

An AI system applied to guidance does not mechanically produce the effects expected of it. Several inputs received during the competition led us to make explicit what we were carrying implicitly: it is the human framework within which the service is inscribed that determines its capacity to effectively reduce inequalities of support, and not the technical performance of the conversational module taken in isolation.

The keynote on Human-Centered AI laid down as a principle that an AI system must be designed with the people it affects, and that it must leave them the possibility of questioning it and of making it evolve. The meta-analysis by Di Pietro and Castaño Muñoz published by the Joint Research Centre (Di Pietro & Castaño Muñoz, 2025) sheds light on this principle on empirical ground. Synthesising 740 estimates from 72 studies, it concludes that access to a digital tool has a positive but modest effect on the academic success of disadvantaged pupils, and that this effect is appreciably greater when the tool is inscribed in a structured pedagogical setting or in active support of families. Several comments received during the competition went in the same direction, recalling that, in matters of guidance, making information available is not sufficient to set in motion the reflective work that an informed choice presupposes.

These contributions reinforced the architecture that places the human counsellor as the final interlocutor, but they also led us to inflect our design of the deployment on two precise points.

The first concerns the methodology of the pilot phases. These will not be able to limit themselves to validating the technical performance of the conversational module. They will have to systematically gather feedback from pupil users and make the product evolve accordingly, which is the practical condition of the principle of design with the people affected. This requirement engages the schedule as much as the method: scaling up will only make sense

if each deployment tier integrates a loop of adjustment to the actual uses observed at the previous tier.

The second concerns the scope of the system. Guidance is not built in the sole face-to-face between a pupil and a tool. It unfolds within a broader ecosystem, which includes parents, *professeurs principaux*, and educational teams, and the meta-analysis cited above shows that it is precisely when this entourage is mobilised that the effect of a digital tool is amplified. The anchoring within AVENIR(s) takes on additional significance here: the platform has, since January 2026, dedicated spaces for parents, and thereby inscribes the system within an overall coherence of the guidance trajectory.

The expected benefit is measured above all in very concrete situations. A parent unaware of the existence of the Dossier Social Étudiant, a *professeur principal* unable to individually examine each Parcoursup file, a National Education psychologist whose interview in practice begins at the fifteenth minute because the context first has to be laid out: in all these configurations, an AI component can free up useful time for the people who matter in a young person's trajectory.

6.3. The conditions for governance over time

Several inputs received during the competition converged towards a single observation: an AI system in guidance cannot be conceived of as a product that is put into service after having been validated once. The nature of the questions it addresses, the sensitivity of its audience, and the rapid evolution of the technologies on which it rests impose a governance designed for the long term. Three angles, epistemic, institutional, and environmental, have specified its contours.

The first angle is that of the cognitive limits inherent to any design team. Alexandra Prigent (CEA-Saclay, AIOLIA project) developed before us the notion of *epistemic gap*: any team that designs an AI system works with a knowledge of risks that is narrower than what would be required to anticipate them, and narrower still than what other disciplines know collectively. No designer, all the more so a two-student team, can secure an ethical system alone. Our logic of anchoring within AVENIR(s) already placed the integration of diverse expertise at the heart of the project, but we envisaged it as an initial moorage to an existing institutional consortium. The keynote led us to make explicit that such an integration only has scope if it is accompanied by a mechanism of audit over time. This requirement converges with a principle that the *OECD Digital Education Outlook 2026* (OECD, 2026) places at the heart of public AI policies in education: the governance of such a system must be adaptive, made up of principles and review mechanisms that evolve with the technology itself, and not of a static framework set once and for all.

The second angle, complementary, is that of institutional sustainability. The remarks of Lucie Jacquet-Malo (Institut Mines-Télécom) underlined that a public AI guidance service could not be envisaged as an isolated initiative, and that it must both anchor itself to public operators whose legitimacy is already recognised, and inscribe itself in financing models compatible with the long duration of an educational system. This twofold requirement, institutional legitimacy and budgetary sustainability, directly intersects with the choice of moorage to AVENIR(s) rather than to an *ad hoc* arrangement, already developed in the previous sub-section, and complements it with a temporal dimension: a service of this nature is not measured by its capacity to be launched, but by its capacity to be maintained over ten or fifteen years under stable conditions of quality and neutrality.

The third angle is that of environmental impact, which did not appear in our initial reasoning in a structured form. The *Référentiel général pour l'IA frugale*, published in the form of an AFNOR Spec by the Ministry of Ecological Transition and the Ecolab of the Commissariat général au développement durable (Ministère de la Transition écologique, 2024), complemented our reasoning on technical choices through this dimension. It first poses a preliminary question that structures everything else: before discussing the sizing of a model or its hosting arrangements, it must be demonstrated that recourse to an AI system is necessary to meet the need, and that no less consumption-intensive solution could meet it as well. This requirement led us to make explicit what existing guidance tools do not enable us to address: databases such as ONISEP or Parcoursup remain relevant for searches structured by keywords, but they do not address the situation in which the pupil does not yet know which terms to mobilise and in which the articulation between programmes and career outcomes presupposes a work of correlation that these databases do not perform.

Two of the good practices ranked by the framework according to their ratio of gain to effort moreover bear directly on our architecture. The mastery of the volume of data converges with our choice of a RAG system relying on official, indexed, and delimited corpora, rather than on the training of a large model from massive datasets. The rationalisation of models validates our orientation towards open-source models of reasonable size, articulated with a specialised re-ranking layer, rather than towards a single large proprietary model. Beyond these architectural choices, we propose to inscribe in OrientAI's governance a commitment to environmental documentation following the life-cycle analysis methodology proposed by the framework, from the pilot phase onwards. This commitment opens, in the longer term, a question that structures the project's posture: a service that, if it succeeds in reducing the frequency of reorientations (each one representing an additional year of schooling with its own footprint), could, on certain categories of impact, fall under the notion of *service à bilan positif* defined by the framework. This hypothesis falls within the scope of long-term evaluation, but it inscribes environmental impact among the indicators of the system's success, and not as a mere constraint to be minimised.

These three angles together sketch a single requirement: OrientAI cannot be thought of as a deliverable validated once and for all. Its epistemic quality, its institutional durability, and its environmental sobriety depend on governance frameworks designed from the outset to evolve with the service, and not on an initial compliance whose maintenance would subsequently be verified.

7. Conclusion

Three movements structure this report.

First, an empirical observation. French post-baccalaureate guidance remains a domain of structural informational inequality, one that the mass adoption of generative AI does not correct, and that the documented biases of these models risk even compounding. The pilot brings to light three convergent mechanisms: the quality of the counselling received varies with the parental education level, the chosen track reflects social origin, and the tracks concentrating reorientations are precisely those towards which the least advantaged pupils orient themselves. These three mechanisms sketch a pattern that generalist LLMs cannot, by construction, either detect or correct.

Second, a proposal. OrientAI is not conceived as a standalone conversational assistant, but as a coherent platform articulating three complementary functions: an AI assistant that informs without prescribing and refuses rather than fabricates, a connection to state-recognised counsellors for the trade-offs that AI cannot carry, and a dashboard that consolidates an administrative trajectory currently scattered. This three-module architecture is not a juxtaposition: it conveys a thesis according to which AI in guidance only makes sense as an information service, integrated into the long timeline of the decision and into the human network that accompanies it.

Third, conditions. The anchoring envisaged within the AVENIR(s) programme (ONISEP) is not an operational detail: it conditions the legitimacy of the service in the eyes of pupils and families, its durability beyond a one-off initiative, and its articulation with state-recognised counsellors. The governance proposed responds to a requirement that this report assumes: an AI system applied to guidance is not measured by its capacity to be launched, but by its capacity to be sustained over ten or fifteen years under stable conditions of quality and neutrality.

The prototype presented here is not intended to constitute a finished solution. Its function is more precise: to demonstrate that a sourced, equitable technical response compatible with European requirements is feasible under modest material conditions. It is now for the public actors concerned to take it up.

8. References

8.1. Career guidance, education, and public policy

Assemblée nationale. (2025, December). *Question écrite n° 11482 sur les effectifs des psychologues de l'Éducation nationale* [Written question No. 11482 on the staffing of National Education psychologists]. <https://questions.assemblee-nationale.fr/q17/17-11482QE.htm>

Bechichi, N. (2026, January). *Quels effets d'une réorientation en première année d'études supérieures sur la réussite des étudiants ?* [What are the effects of reorientation during the first year of higher education on student success?] (IPP Note No. 122). Institut des politiques publiques. <https://www.ipp.eu/publication/quels-effets-dune-reorientation-en-premiere-annee-detudes-superieures-sur-la-reussite-des-etudiants/>

Bourdieu, P., & Passeron, J.-C. (1964). *Les héritiers : les étudiants et la culture* [The inheritors: French students and their relation to culture]. Éditions de Minuit.

Café pédagogique. (2024, November 22). *Orientation : que peut-on faire avec la plateforme Avenirs de l'Onisep ?* [Career guidance: What can be done with Onisep's Avenirs platform?]. <https://www.cafepedagogique.net/2024/11/22/orientation-que-peut-on-faire-avec-la-plateforme-avenirs-de-lonisep/>

Di Pietro, G., & Castaño Muñoz, J. (2025). A meta-analysis on the effect of technology on the achievement of less advantaged students. *Computers & Education*, 226, 105197. <https://doi.org/10.1016/j.compedu.2024.105197>

Ekole. (2025, June). *L'orientation professionnelle arrive trop tôt dans le cursus* [Career guidance comes too early in the curriculum]. <https://www.ekole.fr/blog/lorientation-professionnelle-arrive-trop-tot-dans-le-cursus>

French Government. (2024, June 28). *Programme Avenirs : un projet majeur pour faire évoluer l'accompagnement à l'orientation des jeunes* [Avenirs Programme: A major project to transform career guidance support for young people]. <https://www.info.gouv.fr/actualite/programme-avenirs-un-projet-majeur-pour-faire-evoluer-l-accompagnement-a-l-orientation-des-jeunes-et>

Heaven. (2025). *Born AI 2025 : étude sur les usages de l'IA chez les 18-25 ans* [Born AI 2025: Study on AI usage among 18-25 year olds] [Report]. As cited in Blog du Modérateur. (2025). *93 % des 18-25 ans utilisent l'IA en 2025* [93% of 18-25 year olds use AI in 2025]. <https://www.blogdumoderateur.com/etude-93-18-25-ans-utilisent-ia-2025/>

Inspection générale de l'éducation, du sport et de la recherche. (2020, June). *La réorientation dans l'enseignement supérieur* [Reorientation in higher education] (Report No. 2020-063). Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. <https://www.enseignementsup-recherche.gouv.fr/fr/la-reorientation-dans-l-enseignement-superieur-47369>

Institut des hautes études de l'éducation et de la formation. (2024, March). *Les PsyEN de la spécialité éducation, développement et conseils en orientation scolaire et professionnelle* [PsyEN specialising in education, development, and counselling in academic and career guidance]. <https://www.ih2ef.gouv.fr/les-psyen-de-la-specialite-education-developpement-et-conseils-en-orientation-scolaire-et>

Ministère de l'Éducation nationale. (2022, February 7). *Plan avenir : l'orientation pour que l'avenir soit un choix* [Avenir Plan: Career guidance so that the future is a choice]. <https://www.education.gouv.fr/plan-avenir-l-orientation-pour-que-l-avenir-soit-un-choix-462900>

Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche. (2025, July 2). *Mise en œuvre du plan Avenir à partir de l'année scolaire 2025-2026* [Implementation of the Avenir plan from the 2025-2026 school year onwards] (Service note NOR MENE2519127N). *Bulletin officiel de l'Éducation nationale*, 27. <https://www.education.gouv.fr/bo/2025/Hebdo27/MENE2519127N>

OECD. (2026). *OECD digital education outlook 2026*. OECD Publishing. https://www.oecd.org/en/publications/oecd-digital-education-outlook-2026_062a7394-en.html

OpinionWay, & Edumapper. (2025, June). *Les jeunes et l'orientation scolaire et professionnelle 2025* [Young people and academic and career guidance 2025]. <https://www.opinion-way.com/fr/publications/les-jeunes-et-lorientation-scolaire-et-professionnelle-2025-20373/>

Sénat. (2024-2025). *Amendement n° II-20 au projet de loi de finances* [Amendment No. II-20 to the Finance Bill]. https://www.senat.fr/enseance/2024-2025/143/Amdt_II-20.html

Terrier, C., Hakimov, R., & Schmacker, R. (2023, July). *Confiance en soi et choix d'orientation sur Parcoursup : enseignements d'une intervention randomisée* [Self-confidence and orientation choices on Parcoursup: Findings from a randomised intervention] (IPP Note No. 93). Institut des politiques

publiques. <https://www.ipp.eu/publication/confiance-en-soi-et-choix-dorientation-sur-parcoursup-enseignements-dune-intervention-randomisee/>

8.2. AI, algorithmic bias, and ethical issues in education

Bender, E. M., Gebru, T., McMillan-Major, A., & Mitchell, M. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>

Gupta, A., Patil, N., Ghosh, S., & Gaikwad, S. N. (2026). *Compounding disadvantage: Auditing intersectional bias in LLM-generated explanations across Indian and American STEM education* [Preprint]. arXiv. <https://arxiv.org/abs/2601.14506>

Gupta, S., & Ranjan, R. (2024). *Evaluation of LLMs biases towards elite universities: A persona-based exploration* [Preprint]. arXiv. <https://arxiv.org/abs/2407.12801>

Ministère de la Transition écologique. (2025, June 5). *Publication du référentiel général de l'IA frugale : s'attaquer à l'impact environnemental de l'IA* [Publication of the general framework for frugal AI: Tackling the environmental impact of AI]. <https://www.ecologie.gouv.fr/presse/publication-du-referentiel-general-lia-frugale-sattaquer-limpact-environnemental-lia>

Pôle Léonard de Vinci, & Talan. (2024, May). *Étude sur l'usage de l'IA générative chez les étudiants français* [Study on the use of generative AI among French students]. As cited in Developpez.com. <https://intelligence-artificielle.developpez.com/actu/358003/>

Shailya, Mishra, P., Krishnan, K., & Ravindran, B. (2025). *Where should I study? Biased language models decide!* [Preprint]. arXiv. <https://arxiv.org/abs/2509.04498>

Weka. (2026, January 12). *Albert : l'outil d'IA générative expérimenté à France Services ne sera pas généralisé* [Albert: The generative AI tool piloted at France Services will not be rolled out] [Article based on an AFP dispatch]. <https://www.weka.fr/actualite/administration/article/albert-l-outil-d-ia-generative-experimente-a-france-services-ne-sera-pas-generalise-209194/>

8.3. RAG architecture and LLM techniques

Carbonell, J., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 335-336). Association for Computing Machinery. <https://doi.org/10.1145/290941.291025>

Cormack, G. V., Clarke, C. L. A., & Büttcher, S. (2009). Reciprocal rank fusion outperforms Condorcet and individual rank learning methods. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 758-759). Association for Computing Machinery. <https://doi.org/10.1145/1571941.1572114>

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). *Retrieval-augmented generation for large language models: A survey* [Preprint]. arXiv. <https://arxiv.org/abs/2312.10997>

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474. <https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html>

Mandikal, P., & Mooney, R. (2024). *Sparse meets dense: A hybrid approach to enhance scientific document retrieval* [Preprint]. arXiv. <https://arxiv.org/abs/2401.04055>

Tonmoy, S. M. T. I., Zaman, S. M. M., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A. (2024). *A comprehensive survey of hallucination mitigation techniques in large language models* [Preprint]. arXiv. <https://arxiv.org/abs/2401.01313>

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824-24837. https://proceedings.neurips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html

8.4. Appendix

- OrientAI Platform: <https://orientai-platform.fr>
- OrientAI Questionnaire : https://docs.google.com/forms/d/e/1FAIpQLSey_7AznA5zYms_ww3nbCupgXpdPJC_Aqmq7-liuPtSByBEBw/viewform?usp=header
- GitHub Repository - OrientAI LLM: <https://github.com/matjussu/OrientAI>
- GitHub Repository - OrientAI Platform: https://github.com/matjussu/OrientAI_Platform.git