Hi!Paris Report

Team 11

November 30, 2025

1 Overview and Problem

In many school systems, a large share of students finish compulsory education without solid mastery of core concepts, especially in mathematics. Misunderstandings accumulate over years and often appear only in high-stakes exams, when it is difficult and costly to intervene. Teachers see difficulties in their own classrooms, but their view is local; curriculum designers and ministries mostly rely on aggregate exam scores and occasional evaluations, which do not show which concepts are misunderstood, how, and by whom.

Support is also unequally distributed. Some students have access to parents, tutors and curated online resources; others do not. Generic AI chatbots are misaligned with national curricula, can hallucinate, and provide no usable feedback to the education system. At the same time, employers have little direct evidence of students' actual competences and learning habits, and hiring depends heavily on institution brand, networks and generic tests.

Curriculum Compass addresses these inefficiencies. It uses a curriculum-aligned large language model (LLM) to provide real-time support to students and teachers while generating privacy-preserving data that can guide curriculum design, vocational pathways and research.

Team Information

Team Number: 11
Members:

- Ayman Orkhis: IP Paris
- Basar Inan: HEC Paris
- Belhaj Othmane: IP Paris
- Boudelaa Fares: IP Paris
- Osman Bagci: IP Paris

2 Solution: Curriculum Compass

Curriculum Compass is a national platform built around an LLM specialised on the official curriculum and grounded on ministry-approved content. All answers are constrained to this corpus, ensuring consistency with what teachers teach and exams assess.

The Student Companion is a conversational assistant accessible via web, mobile or existing learning environments. Students ask questions in natural language, and the assistant returns explanations and examples that follow the syllabus, at a level of abstraction adapted to age and prior knowledge. The aim is to make reliable clarification available to every student between lessons, without replacing teachers.

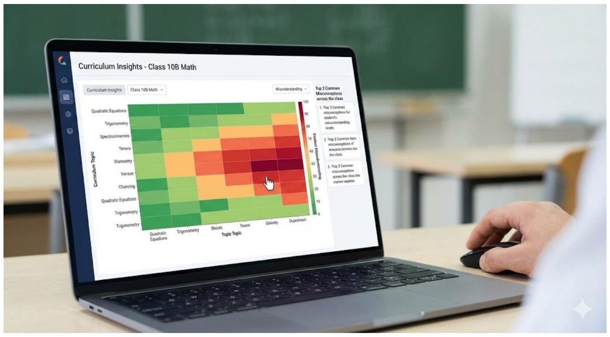

Figure 1: Example of class-level Curriculum Insights heatmap (misunderstandings by topic).

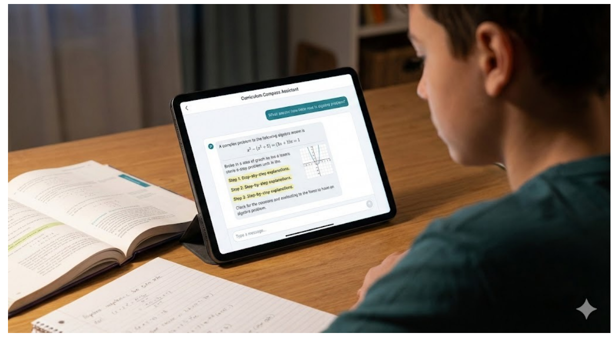

Figure 2: Example of AI tutor helping students on course content

Curriculum Insights turns every interaction into structured data. Questions are anonymised and tagged by subject, concept, type of misunderstanding, grade, school and time. Aggregated dashboards then show where confusion concentrates and how patterns change after curriculum or textbook revisions. The same system that tutors individual students thus becomes a continuous sensor of conceptual understanding for teachers, school leaders and curriculum architects.

Once the core platform is established, a vocational layer can be added. Corporate partners design voluntary training modules hosted on Curriculum Compass. Students who opt in complete sector-specific tasks; employers gain a richer view of engagement and competence than a CV alone permits, under the oversight of education authorities.

3 Value Proposition

For students, Curriculum Compass offers immediate, curriculum-aligned explanations regardless of background or location. Misunderstandings are addressed earlier, supporting confidence and persistence, and AI is used in a guided, school-governed context.

For teachers, the platform reduces repetitive clarification work and provides a clear view of class-level difficulties. Instead of relying on intuition, they see which concepts generate questions and which misconceptions recur, and can adjust pacing and remediation accordingly.

For curriculum designers and ministries, Curriculum Compass creates a continuous feedback loop on how the curriculum is experienced in classrooms. They can monitor how different groups engage with specific concepts and how reforms affect the pattern of questions, and can adapt curriculum, assessment and teacher training on the basis of current data rather than infrequent exam cycles.

For employers, the vocational layer enables talent discovery based on observable learning behaviour and performance in realistic tasks, not only on credentials and networks. Students gain structured exposure to sectors and can signal interest and ability through engagement with company modules.

For researchers and society, anonymised longitudinal data on learning processes and outcomes allow rigorous study of learning dynamics, inequalities and long-term trajectories, informing both education and labour-market policy.

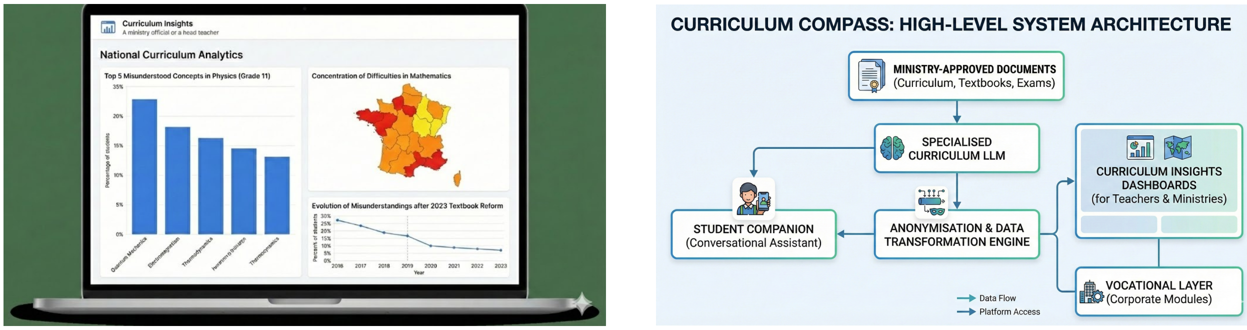

Figure 3: Country-level analytics (left) and **system architecture view (right).**

4 Business Model and Revenue Streams

Curriculum Compass is impact-driven but designed to be financially sustainable through three complementary revenue streams.

The core stream is a B2G subscription model. Ministries of education and regional authorities pay an annual fee, per student or per school, covering operation of the LLM and infrastructure, ingestion and updating of curriculum materials, analytics access, and training and support. Initial adaptation—digitising curriculum, fine-tuning the model, building dashboards—can be financed via setup fees or spread over early subscription years. Using efficient, domain-specific models and retrieval-augmented generation keeps compute needs moderate and per-student costs comparable to a small share of existing textbook or digital budgets.

A second stream comes from corporate vocational partnerships. Once the platform has proven educational value, companies pay to host their vocational modules on Curriculum Compass. Fees cover a secure environment for sector-specific content, aggregated analytics on participation and performance, and ministry-approved mechanisms for identifying students who have voluntarily engaged with their learning paths. The core educational service remains publicly governed and free at the point of use.

The third stream arises from governed data partnerships. The platform generates large volumes of de-identified, aggregated data on learning patterns and conceptual mastery. Under strict frameworks agreed with education authorities, selected EdTech companies can license access to specific indicators or APIs to improve their products, and academic institutions can pay access fees, typically via research grants, to work with anonymised datasets or run experiments. Strong rules on anonymisation and ethical use ensure that monetisation remains compatible with privacy and public trust.

Together, these streams diversify financial risk, reduce dependence on a single public budget and align incentives across authorities, employers, innovators and researchers.

5 Implementation, Integration and Risks

Adoption at scale requires smooth integration into existing digital and organisational routines.

In a pilot stage, the platform is deployed in a small but diverse set of schools, initially in one subject and a limited range of grades. It connects to current digital workspaces via APIs and single sign-on, teachers receive concise training and feedback is collected on answer quality, usability, dashboard clarity and technical robustness. This stage is used to refine features and governance.

In a gradual rollout, the ministry extends the platform across regions and subjects, with particular attention to disadvantaged areas and key transition years. Online training and documentation support educators; communication campaigns explain objectives and privacy guarantees to families. Governance structures are formalised, including a steering committee with representatives from the ministry, teachers, students and independent experts.

In an extension stage, vocational modules and data partnerships are activated under clear rules. Corporate content is introduced gradually and evaluated for educational value and equity. Procedures for research and EdTech access to anonymised data are transparent, with defined approval workflows and audits. Across all stages, success is monitored through usage, changes in concept-level misunderstanding patterns, teacher feedback and early evidence on exam performance and equity.

Key risks include over-reliance on AI, privacy concerns, unequal access linked to the digital divide and institutional resistance. Mitigation involves presenting the assistant as a complement rather than a substitute for teaching, anonymising and aggregating data before analysis, prioritising under-served regions and low-bandwidth solutions in rollout, and involving teachers and their representatives in design and governance from the outset.

6 Scientific Approach

6.1 Technical Foundation

Our solution employs a machine learning pipeline designed to predict student performance while maintaining computational efficiency and scalability.

6.1.1 Pre-Processing Strategy

We implemented a rigorous data cleaning approach to ensure model quality and prevent overfitting:

- **Missing Data Handling**: We identified and removed features with more than 80% missing values, as these would contribute primarily noise rather than predictive signal to our models.
- **Data Leakage Prevention**: We deliberately excluded "Math-qXX-average-score" features from our training data, as these variables would create unrealistic performance metrics by leaking information about the target variable.

6.1.2 Feature Engineering

To reduce dimensionality and improve model interpretability, we consolidated redundant features that conveyed similar information. This approach not only decreased computational requirements but also helped mitigate the risk of multicollinearity in our models.

6.1.3 Model Selection and Optimization

Our data is tabular we thus test multiple gradient boosting models to identify the optimal solution:

- **XGBoost**: Selected as our primary model due to its flexibility and superior performance on our validation set
- **LightGBM**: Evaluated for its speed advantages on large datasets
- **CatBoost**: Tested for its native handling of categorical variables
- **Hybrid Model (XGBoost-LightGBM-CatBoost)**: Explored to potentially combine the strengths of individual models

We optimized model performance through two complementary hyperparameter tuning strategies:

- **GridSearchCV**: For exhaustive search over specified parameter grids
- **Optuna**: For more efficient Bayesian optimization of the hyperparameter space

6.2 Future Development

To further improve our solution, we have identified several enhancement opportunities:

- **Data Standardization**: Implementing feature scaling to normalize value ranges across different features
- **Advanced Imputation**: Developing domain-specific strategies for handling missing values rather than simple deletion, potentially recovering valuable signal from partially complete records

7 Conclusion

Curriculum Compass turns AI from a stand-alone tutoring tool into system-level infrastructure for more equitable and data-driven education. By aligning an LLM with the national curriculum and treating student questions as a continuous diagnostic signal, it helps learners resolve misunderstandings earlier, gives teachers actionable insight into class difficulties and enables ministries to adjust curricula and teacher training using current evidence.

Through its vocational and data extensions, the platform also connects education systems to labour markets and the research community, addressing long-standing inefficiencies in talent discovery and in understanding what works in education. Supported by a diversified, impact-aligned business model and careful governance, Curriculum Compass can improve learning outcomes, reduce inequality and better prepare students for the opportunities and challenges of the AI age.