Public Report

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

1. Integrantes do time de projeto

Nome completo
Antônio Ribeiro Cavalcante
Yasmin Vitória Rocha de Jesus

2. Introduction

The third module of the *Ensigna* project represents a significant milestone in its technical and scientific evolution. While the previous stages were dedicated to understanding the problem, defining hypotheses, and validating the initial product concept, this module focused on consolidating the platform's technological foundation and preparing it for scalability. The primary goal was to transition from theoretical design to a fully functional system, connecting artificial intelligence, computer vision, and natural language processing into an integrated and consistent experience.

At this stage, Ensigna evolved from a prototype into a more mature educational platform capable of recognizing Libras gestures, providing real-time feedback, and transforming learning into an interactive and inclusive experience. The solution combines a robust backend infrastructure based on microservices, a clean and modular architecture, and specialized components such as the Chatbot Service for natural language processing and the Computer Vision Service for gesture recognition. Each of these elements contributes to creating a cohesive ecosystem that ensures high performance, security, and usability.

Beyond the technical scope, this module also addressed strategic aspects of intellectual property protection, business scalability, and financial sustainability. The Ensigna project reinforced its alignment with the social responsibility and inclusion objectives that guide its conception, consolidating itself as an innovative educational tool that integrates technology, accessibility, and ethical design. The progress made in this module thus represents not only a technical advance but also a step toward transforming how sign language education is delivered in digital environments.

3. Product Thesis and Hypothesis Validation

The second section of this report focuses on the process of validating the product thesis and the main hypotheses that guided the development of Ensigna throughout this module. From the beginning, the project was conceived not only as a technological innovation but also as a response to real social and corporate demands related to accessibility and inclusion. Therefore, the hypotheses were structured to bridge the gap between business needs, user expectations, and educational effectiveness.

The first hypothesis established that companies demonstrate a growing interest in promoting inclusion and accessibility within their teams but often lack adequate tools or clear strategies to do so. Through Ensigna, it was possible to verify the potential for scalable, data-oriented Libras training that meets corporate requirements while fostering a culture of inclusion. This insight came from qualitative analyses and conversations with professionals from different sectors, confirming that accessibility initiatives need measurable, continuous, and cost-efficient solutions.

The second hypothesis proposed that users learn more effectively when they receive immediate, visual, and interactive feedback during the learning process. The integration of artificial intelligence and computer vision technologies in the Ensigna platform allowed this hypothesis to be validated through internal experiments. These tests demonstrated that real-time gesture recognition significantly enhances user engagement and learning retention, confirming the pedagogical potential of real-time feedback mechanisms.

Finally, the third hypothesis highlighted that traditional Libras learning methods—mainly face-to-face courses or recorded content, are expensive, limited in reach, and often demotivating. The adoption of a digital, interactive, and flexible model not only reduces costs but also democratizes access to training. The digital format, combined with gamification strategies, proved capable of maintaining user motivation over longer learning periods.

The validation of these hypotheses consolidated Ensigna as a platform that transcends the boundaries of conventional teaching methods. It positions itself as an innovative solution that unites technological accuracy, social purpose, and practical impact. By confirming the technical feasibility and pedagogical effectiveness of its approach, Ensigna advances toward becoming a reference model in inclusive education, offering both individuals and companies a tool that combines engagement, scalability, and social responsibility.

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

4. MVP Development and Validation

The development of the Minimum Viable Product (MVP) represented the critical transition from conceptual validation to tangible experimentation. In this phase, the Ensigna project was transformed from a theoretical proposition into an operational platform capable of validating its pedagogical and technical premises in practice. The MVP was designed to recognize and classify gestures from the Libras alphabet through a computer vision model based on convolutional neural networks (CNN) combined with long short-term memory (LSTM) architecture, ensuring the interpretation of temporal sequences of hand movements.

The MVP's implementation followed an iterative methodology, integrating software engineering best practices with machine learning experimentation cycles. The model was trained using a controlled dataset of alphabet gestures, developed through the Image Capture System, which ensured image consistency and diversity. After training, the system was connected to a backend built with FastAPI, responsible for handling requests, running real-time inference, and providing structured outputs to the frontend interface.

During the validation process, internal tests confirmed the system's ability to recognize letters with a high degree of accuracy and stability. The results demonstrated not only the technical viability of real-time detection but also the usability and responsiveness necessary for an engaging learning experience. Each inference was accompanied by visual and textual feedback to the user, simulating the environment of a real learning session.

In addition to the model's performance, particular attention was given to optimizing inference speed and ensuring synchronization between the backend and the user interface. The experiments confirmed that the architecture could maintain real-time performance with minimal latency, enabling an immersive experience that supports the user's cognitive learning process.

Thus, the MVP validation phase not only confirmed the robustness of the model but also established the functional foundation for future iterations of the platform. The results consolidated Ensigna's potential to evolve into a scalable, educational, and interactive tool that applies artificial intelligence to strengthen accessibility and inclusion in both academic and corporate environments.

5. Technical Architecture

The architecture of the Ensigna platform was designed to ensure modularity, scalability, and robustness, following contemporary paradigms of distributed systems and clean architecture principles. This structural foundation enables each service to evolve independently while maintaining cohesion through well-defined communication protocols. The modular composition of Ensigna encompasses distinct layers: Frontend, Backend Gateway, Chatbot Service, Image Search, and Computer Vision, ach with specific responsibilities and interfaces.

At the core of the system, the Backend Gateway acts as the main orchestrator, mediating interactions between users and the microservices that compose the ecosystem. It manages authentication, validation, caching, and rate limiting, ensuring that the requests are efficiently processed and routed to the appropriate services. Communication between modules is established through REST APIs and WebSocket channels, allowing both synchronous and asynchronous operations to occur seamlessly.

The Chatbot Service processes linguistic inputs, normalizing user messages and structuring them into semantic representations that the system can interpret. This component serves as the linguistic engine of the platform, enabling the association between textual inputs and visual feedback. Meanwhile, the Image Search Service retrieves the appropriate visual representations of Libras gestures based on the processed text, bridging the semantic and visual dimensions of the user experience.

The Computer Vision Service constitutes the technological backbone of the system. It performs real-time gesture recognition through a three-stage processing pipeline: landmark detection via MediaPipe, temporal sequence modeling through LSTM neural networks, and classification of hand gestures. This service exposes REST endpoints for frame streaming, capture control, and prediction retrieval. It operates under optimized conditions for low latency and high throughput, ensuring consistent feedback in real time.

The Frontend layer provides the user interface for all system interactions. Developed to be accessible and intuitive, it handles both text inputs and gesture feedback visualization. The frontend communicates directly with the backend through HTTP and WebSocket channels, receiving predictions, annotated frames, and interaction metrics.

Underneath these functional layers lies the Data Layer, composed of a relational database for user and session management, an object storage system for multimedia content and model artifacts, and an in-memory cache for transient data. The architecture also incorporates observability mechanisms such as structured logging, telemetry, and distributed tracing, allowing continuous monitoring of performance and reliability metrics.

From a deployment perspective, the system operates through containerized services orchestrated via Docker Compose for development and Kubernetes for production environments. This design supports independent horizontal scaling, ensuring that each service can adapt to variable workloads. Furthermore, the architecture adheres to LGPD data protection principles, incorporating access control policies, secure communication channels, and controlled data retention mechanisms.

In summary, Ensigna's architecture integrates software engineering rigor with machine learning innovation, resulting in a system capable of sustaining real-time educational experiences with efficiency and reliability. The modular design not only guarantees technical stability but also allows the platform to evolve continuously, accommodating new pedagogical functionalities and AI enhancements without compromising performance or security.

6. Chatbot Service (NLP Layer)

The Chatbot Service is a fundamental microservice within the Ensigna ecosystem, responsible for linguistic normalization and semantic preparation of user inputs. Its function extends beyond simple text processing; it acts as an intermediary between natural language and the visual learning experience, ensuring that every user message is interpreted consistently before being transformed into visual and gestural outputs.

From an architectural perspective, the Chatbot Service follows the principles of Clean Architecture and Model–View–Controller (MVC) design. This structure allows for a clear separation of concerns between the linguistic logic, data handling, and interface exposure layers. Developed in Python 3.12 with FastAPI, the service leverages SpaCy and NLTK libraries for natural language processing, tokenization, and lemmatization. Its communication with the Backend Gateway occurs through RESTful endpoints using the HTTP POST method and JSON payloads.

The service receives user messages, applies normalization procedures, and returns a structured JSON containing linguistic metadata, such as tokenized words, letter counts, and detailed breakdowns of the text. This output enables the integration between text semantics and image-based sign language feedback, supporting the educational flow of the Ensigna platform.

The main endpoint of the service is defined as POST /chatbot/message, which expects a JSON request containing the text message and language parameter. The service performs several internal stages: it removes punctuation and accents, standardizes case, applies lemmatization to reduce verbs to their infinitive form, and finally generates token and character statistics. These transformations ensure that every input is handled uniformly, regardless of the user's writing style or linguistic variations.

Example of Request and Response:

```
Request:
{
  "message": "Eu amei comer isso.",
  "lang": "pt-BR"
}

Response:
{
  "status": "ok",
  "statusCode": 200,
  "normalization": {
    "neutralText": "eu amar comer isso",
    "tokens": ["eu", "amar", "comer", "isso"],
    "lengthWords": 4,
    "lengthLetters": 14,
    "lettersByToken": [
      ["e","u"],
      ["a","m","a","r"],
      ["c","o","m","e","r"],
      ["i","s","s","o"]
    ]
  }
}
```

This example illustrates the systematic transformation of a natural language sentence into a normalized structure ready for downstream processing. The service outputs a neutral text

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

version, lists individual tokens, calculates statistical properties, and decomposes each word into its letters, facilitating the mapping to visual elements within the system.

To ensure robustness and reliability, the Chatbot Service incorporates comprehensive error handling and logging. Possible responses include HTTP codes 400 for invalid payloads, 415 for unsupported media types, and 500 for unexpected internal errors. All exceptions are logged using structured JSON logs, allowing efficient debugging and performance monitoring.

In addition to its technical relevance, this microservice plays a key pedagogical role in maintaining the fluidity and contextual accuracy of the user's learning journey. By transforming user input into a structured and analyzable form, the Chatbot Service bridges linguistic processing and visual interpretation, ensuring that the Ensigna platform delivers a coherent and adaptive educational experience in real time.

7. Computer Vision and Backend Model

The Computer Vision and Backend Model represent the technological foundation that enables Ensigna's core functionality: the real-time recognition of Libras gestures. This component was meticulously developed to ensure accuracy, efficiency, and scalability, integrating advanced artificial intelligence techniques within a service-oriented software architecture.

The backend employs a modular design built with FastAPI, where each route exposes a specific functionality related to gesture detection and prediction. The primary objective of this system is to process image sequences from a camera feed, extract relevant features, and classify hand gestures according to the Libras alphabet. This process is made possible through a combination of MediaPipe, used for hand landmark detection, and TensorFlow/Keras, responsible for training and executing an LSTM (Long Short-Term Memory) neural network capable of modeling temporal dependencies across sequential frames.

The complete processing pipeline follows a structured flow. First, the camera feed is captured through OpenCV, and each frame is analyzed by the MediaPipe model, which identifies 21 three-dimensional landmarks per hand. These landmarks are then organized into numerical vectors that form a temporal sequence used by the LSTM classifier. The model outputs the most probable letter label and a confidence score, which are then sent to the API layer for real-time feedback.

The system exposes several key endpoints that form the interaction interface with external applications. The `/start_detect` and `/stop` routes control the initiation and termination of the capture process, while `/stream` provides a continuous MJPEG or WebSocket stream of

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

annotated frames. The `/prediction` endpoint delivers the most recent inference result, encapsulated in a structured JSON response, as shown below:

Example of Response:


```json
{
  "status": "ok",
  "prediction": {
    "label": "B",
    "confidence": 0.87,
    "timestamp": 1694442000.123
  },
  "message": "Letter 'B' identified with 87% confidence"
}
```


This format allows the frontend to visualize predictions immediately, enabling users to receive real-time feedback on the accuracy of their gestures. The response includes both quantitative (confidence level) and qualitative (semantic interpretation) data, ensuring transparency and interpretability.

From a software engineering standpoint, the backend integrates multi-threading and efficient frame buffering to maintain stability and minimize latency, achieving near real-time performance at 30 FPS. Additionally, structured error handling mechanisms prevent service interruptions, while cross-origin resource sharing (CORS) middleware ensures interoperability with multiple frontend clients.

The combination of a neural model optimized for sequential gesture analysis and a lightweight, high-performance API layer demonstrates Ensigna's capacity to handle the computational challenges inherent to computer vision applications. The system is capable of processing continuous image streams, performing inference in milliseconds, and providing outputs that preserve the pedagogical intent of immediate and interactive learning.

Overall, the Computer Vision and Backend Model together embody the core of Ensigna's intelligent architecture. They transform visual data into actionable educational feedback, bridging the gap between human communication and artificial intelligence, and reinforcing the platform's mission to promote inclusion through technology.

8. Image Capture System

The Image Capture System is a key component in the data acquisition pipeline of Ensigna, responsible for generating a standardized and high-quality dataset used to train and validate the computer vision models. Its creation was guided by the need for consistency, diversity, and control over the images representing the Libras alphabet, ensuring that the learning model generalizes effectively across different users and environmental conditions.

Developed in Python, the system integrates Tkinter for graphical interface management and OpenCV for camera access and frame processing. It allows users to record sequences of gestures from A to Z in a structured and automated manner. The interface was designed with usability in mind: users select the desired letter, camera, and storage directory, after which the system initiates a pre-capture countdown followed by automatic image recording.

The data acquisition workflow was optimized to balance simplicity and control. Each recording session begins with a five-second countdown displayed on the screen, helping the participant position their hand correctly within the capture area. Afterward, the system records for five seconds at an approximate rate of five frames per second, resulting in about twenty-five images per session. The frames are automatically saved in directories labeled by letter (A–Z), with file names encoded with timestamps for traceability. This structure facilitates later processing and labeling during model training.

To ensure dataset quality, the system follows a set of best practices in image collection. The environment is controlled to maintain neutral backgrounds and adequate lighting, while framing guidelines help prevent occlusion or loss of important gesture details. The software encourages consistency across sessions and participants, while also promoting diversity through the inclusion of different skin tones, hand sizes, and orientations. This balance between uniformity and variability strengthens the robustness of the trained models.

From a technical architecture standpoint, the system is divided into four primary layers: Presentation, Application, Domain, and Infrastructure. The presentation layer handles user interaction through the GUI; the application layer orchestrates capture and timing logic; the domain layer manages gesture sequence and preprocessing; and the infrastructure layer deals with video access, file storage, and metadata registration. This layered organization promotes maintainability and scalability, allowing easy addition of new functionalities such as video capture or cloud synchronization.

The Image Capture System also supports metadata generation for each session, including parameters like camera index, frame rate, duration, timestamp, and user identification. These records are stored in JSON format alongside the images, contributing to reproducibility and dataset governance. The system's modular design allows horizontal scaling—multiple stations can capture data simultaneously and later synchronize outputs to a central repository, such as an S3 or GCS bucket.

Beyond its technical utility, this system plays an essential role in Ensigna's ethical and inclusive mission. By structuring the dataset under principles of diversity, transparency, and consent, it aligns with responsible AI development practices. The result is a robust dataset that serves as the empirical foundation for accurate gesture recognition models, directly supporting the pedagogical and social goals of the platform.

9. Patent and Intellectual Property

The protection of intellectual property (IP) within the Ensigna project is a strategic and structural element that ensures both technological credibility and long-term sustainability. This dimension of the project reflects an understanding that innovation in educational technology requires not only technical excellence but also legal safeguarding of creative assets. Consequently, the IP strategy developed in Module 3 aims to protect the brand, software, and interface design, consolidating Ensigna's unique position in the intersection of accessibility, education, and artificial intelligence.

The IP framework was organized into three main protection layers, each targeting a specific aspect of the solution. The first layer involves the Trademark Registration, which guarantees the exclusive rights to use the name *Ensigna* in association with its educational, technological, and social purposes. The registration process follows the Nice Classification, with Classes 41 (Education and Training), 42 (Software as a Service and Technology Development), and 35 (Corporate Social Responsibility and Business Services). These classifications ensure broad coverage for national and international operations, supporting future scalability and brand expansion.

The second layer refers to the Industrial Design Registration, which covers the visual identity of the platform's interface. This registration aims to protect the aesthetic elements of the user experience—such as layout, color patterns, and component arrangements—ensuring that Ensigna's distinctive and inclusive visual design cannot be replicated or modified without authorization. The interface was designed not only for usability but also for accessibility, incorporating contrast standards and assistive design principles to accommodate users with different visual and motor needs.

The third protection layer concerns the Computer Program Registration, focused on safeguarding the platform's source code and technical architecture. This registration certifies the originality and authorship of the developed software, particularly its backend

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

microservices, computer vision algorithms, and natural language processing modules. By protecting these assets, the registration prevents unauthorized replication or commercial exploitation of Ensigna's proprietary technology. The inclusion of code documentation and versioning history in the registration further supports authenticity and traceability.

The comprehensive IP strategy adopted by Ensigna extends beyond legal compliance; it represents a deliberate step toward consolidating the project's identity as a scientific and entrepreneurial initiative. Intellectual property rights reinforce investor confidence, facilitate potential partnerships, and establish the groundwork for international expansion through mechanisms such as the Madrid Protocol. Moreover, the integration of ethical and transparent practices into the registration process aligns with Ensigna's commitment to social responsibility and respect for creative work.

In summary, the intellectual property protection of Ensigna encompasses the safeguarding of its brand, interface, and underlying technology. This triad of protection mechanisms ensures that the platform's innovations remain exclusive and that its value as a socially transformative solution is preserved. Beyond ensuring legal ownership, this strategy reinforces the academic and professional legitimacy of Ensigna, supporting its trajectory toward becoming a reference model in inclusive technological education.

10. Growth Hacking and Investment Strategy

The growth and investment strategy defined in Module 3 represents a pivotal stage in the transition of Ensigna from a research-based initiative to a viable and scalable business model. This component integrates entrepreneurial planning, financial modeling, and digital marketing methodologies under the same strategic framework, with the objective of ensuring that Ensigna achieves sustainable expansion while maintaining its educational and social mission.

The growth strategy is grounded in the AAARRR funnel model (Acquisition, Activation, Retention, Referral, Revenue, and Reactivation), which allows for systematic tracking of user engagement throughout the platform's lifecycle. Each stage of the funnel was designed to optimize both pedagogical impact and commercial traction. In the Acquisition phase, partnerships with universities, inclusion-focused organizations, and corporate training programs serve as the main entry points for new users. These alliances not only promote the dissemination of Ensigna but also reinforce its institutional credibility. The Activation stage focuses on the user's first interaction with the system, emphasizing gamified learning modules and instant visual feedback to foster engagement and reduce initial dropout rates.

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

The Retention phase prioritizes the user's continuous interaction through personalized dashboards, progress-tracking systems, and performance analytics that provide immediate feedback on learning outcomes. These features strengthen user motivation and promote long-term participation. The Referral mechanism leverages the network effect by encouraging companies and educational institutions to recommend the platform as part of their accessibility and diversity programs. Finally, the Revenue component is sustained through a B2B SaaS (Software as a Service) model, in which clients subscribe to different tiers based on user volume, training hours, and additional functionalities such as customized dashboards and extended support. This structure allows for flexible pricing while maintaining predictable and recurring revenue streams.

The investment strategy complements this growth model by focusing on financial stability and scalability. The financial plan estimates an initial CAPEX of approximately R$13,000, encompassing infrastructure development, cloud service provisioning, and intellectual property registration. The OPEX (Operational Expenditure) is projected at R$3,500 per month, which sustains the platform for up to 250 active users under current infrastructure conditions. With this structure, the break-even point is expected to occur once Ensigna secures at least five corporate clients under standard subscription plans. This calculation accounts for both fixed and variable costs, allowing the project to maintain financial equilibrium while reinvesting in model improvement and user experience.

To attract early-stage investors, Ensigna positions itself as a hybrid impact-driven startup that combines social transformation with technological innovation. Its dual focus—educational accessibility and artificial intelligence—places it within two rapidly growing markets, enhancing its appeal for angel investors and ESG-aligned funds. Furthermore, the platform's intellectual property protection and validated MVP increase investor confidence by reducing operational and technological risks. Future funding rounds are expected to prioritize infrastructure expansion, the incorporation of MLOps pipelines, and international scalability.

Beyond its financial dimensions, the growth strategy also encompasses marketing and communication initiatives centered on growth hacking techniques. These include content marketing campaigns that emphasize inclusion through technology, the development of ambassador programs with accessibility advocates, and data-driven experimentation to optimize user acquisition costs (CAC) and maximize lifetime value (LTV). Continuous performance metrics will be tracked using tools such as Google Analytics and CRM integrations to refine conversion funnels and identify potential growth bottlenecks.

In conclusion, the Growth Hacking and Investment Strategy consolidates Ensigna's trajectory as both a sustainable enterprise and a socially transformative platform. Through the combination of sound financial planning, adaptive marketing strategies, and a commitment to inclusivity, the project establishes a solid foundation for long-term expansion. This strategy not only ensures the economic viability of Ensigna but also strengthens its positioning as a benchmark in the field of inclusive educational technology.

11. Conclusions and Next Steps

The conclusion of Module 3 represents a defining moment in the evolution of the Ensigna project, marking the transition from theoretical experimentation to a structured and functional technological ecosystem. Over the course of this stage, the project solidified its scientific, technical, and strategic dimensions, bringing together artificial intelligence, computer vision, and linguistic processing under a coherent and scalable architecture. The advancements achieved consolidate Ensigna as a platform capable of delivering measurable social impact while maintaining the robustness required for long-term operation.

From a technical standpoint, the completion of the MVP and the integration of modular microservices demonstrate that the platform has reached a state of maturity. The components developed—such as the Chatbot Service, the Computer Vision model, and the Image Capture System—interconnect seamlessly within a clean architectural framework, ensuring performance, reliability, and maintainability. These results validate the system's capacity to recognize gestures in real time, normalize linguistic inputs, and generate educational feedback through a fully functional interface.

On a strategic level, the intellectual property protection, growth plan, and investment roadmap developed during this module establish the foundation for Ensigna's expansion into the market. By combining financial sustainability with ethical and inclusive objectives, the project aligns technological innovation with social purpose, strengthening its credibility among stakeholders, investors, and end users. The integration of responsible AI practices and adherence to LGPD guidelines further reinforce its readiness for real-world deployment in corporate and academic contexts.

Looking forward, several key directions guide the project's continued development. The next step involves the implementation of MLOps pipelines to automate model training, versioning, and monitoring, ensuring continuous improvement in prediction accuracy and efficiency. Simultaneously, user testing with partner companies and accessibility organizations will provide empirical feedback to refine the user experience and pedagogical design. The deployment of the infrastructure in cloud environments such as AWS or GCP will allow for horizontal scalability and integration with advanced monitoring and analytics tools. Finally, the expansion of the dataset to include complete Libras words and phrases will enhance the model's generalization capabilities, moving the system beyond alphabet recognition toward full linguistic comprehension.

In summary, Module 3 positions Ensigna not only as a technological achievement but also as a socially transformative initiative. The work accomplished throughout this phase demonstrates that inclusive education can be effectively supported by artificial intelligence when guided by principles of accessibility, ethics, and human-centered design. The continuity of this trajectory will enable Ensigna to become a reference point in the field of inclusive educational technology, bridging communication gaps and empowering institutions to build a more accessible and equitable future through innovation.

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br