





## Article

# INTU-AI: Digitalization of Police Interrogation Supported by Artificial Intelligence

José Pinto Garcia <sup>1</sup>, Carlos Grilo <sup>1,2,\*</sup>, Patrício Domingues <sup>1,3</sup> and Rolando Miragaia <sup>1,2</sup>

<sup>1</sup> School of Technology and Management, Polytechnic University of Leiria, 2411-901 Leiria, Portugal; 2232565@my.ipleiria.pt (J.P.G.); patricio.domingues@ipleiria.pt (P.D.); rolando.miragaia@ipleiria.pt (R.M.)

<sup>2</sup> Computer Science and Communication Research Centre, 2411-901 Leiria, Portugal

<sup>3</sup> Instituto de Telecomunicações, 2411-901 Leiria, Portugal

\* Correspondence: carlos.grilo@ipleiria.pt

## Abstract

Traditional police interrogation processes remain largely time-consuming and reliant on substantial human effort for both analysis and documentation. Intuition Artificial Intelligence (INTU-AI) is a Windows application designed to digitalize the administrative workflow associated with police interrogations, while enhancing procedural efficiency through the integration of AI-driven emotion recognition models. The system employs a multimodal approach that captures and analyzes emotional states using three primary vectors: Facial Expression Recognition (FER), Speech Emotion Recognition (SER), and Text-based Emotion Analysis (TEA). This triangulated methodology aims to identify emotional inconsistencies and detect potential suppression or concealment of affective responses by interviewees. INTU-AI serves as a decision-support tool rather than a replacement for human judgment. By automating bureaucratic tasks, it allows investigators to focus on critical aspects of the interrogation process. The system was validated in practical training sessions with inspectors and with a 12-question questionnaire. The results indicate a strong acceptance of the system in terms of its usability, existing functionalities, practical utility of the program, user experience, and open-ended qualitative responses.

**Keywords:** artificial intelligence; police interrogation; multimodal emotion recognition; decision support system



Academic Editor: George Drosatos

Received: 8 August 2025

Revised: 11 September 2025

Accepted: 26 September 2025

Published: 7 October 2025

**Citation:** Garcia, J.P.; Grilo, C.; Domingues, P.; Miragaia, R. INTU-AI: Digitalization of Police Interrogation Supported by Artificial Intelligence. *Appl. Sci.* **2025**, *15*, 10781. <https://doi.org/10.3390/app151910781>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

We are currently experiencing a profound transformation driven by the rapid advancement of digital technologies [1,2]. In the justice sector specifically, there has been a gradual integration of computational tools aimed not only at streamlining administrative procedures but also at supporting decision-making in highly sensitive contexts, such as criminal investigations [3].

Interrogation remains a crucial component in various types of investigations, including civil, workplace, public inquiries, regulatory, and criminal cases. In interrogation scenarios, the analysis of behavior, especially emotions, such as facial expression, voice tone, and others, is crucial for interrogators to assess the credibility and honesty of the questioned person [4]. Leveraging AI models to analyze data, including facial expressions, voice inflections, and text enables the development of innovative tools that aid in emotional assessment within interrogation contexts, thus improving the precision and efficacy of legal processes [5].

In Portugal, the domain of specialized criminal police is divided between two entities: the Polícia Judiciária (PJ) [6] which handles civilian matters, and the Polícia Judiciária Militar (PJM) [7] which oversees military cases. Currently, the interrogation procedure in a criminal investigation police department is a predominantly manual, time-consuming, and demanding process in terms of human effort for both analysis and documentation. In this context, it is crucial to equip the security and justice sector, particularly criminal investigation police forces, with tools capable of assisting interrogators in analyzing interrogation content while also enabling the full digitalization of the associated administrative processes. For this purpose, we developed the INTU-AI software tool. This software aims to streamline the entire administrative workflow within a single platform, while also interrogation procedures through the integration of AI models capable of classifying emotions during interviews. This constitutes a decision-support tool that can complement and improve investigative outcomes.

The main contributions of this work are as follows: (i) Study and applications of algorithms, namely Facial Expression Recognition (FER), Speech Emotion Recognition (SER), and Text-based Emotion Analysis (TEA), to multimodal input sources from interrogation; (ii) Creation of the INTU-AI application to automate sentiment analysis of PJM's interrogations, namely the identification of emotional inconsistencies, as well as to streamline administrative tasks.

The remainder of this paper is organized as follows. Section 2 reviews related work in AI-driven deception and emotion detection. Section 3 addresses the interrogation process and the need for its digitalization. Section 4 focuses on the challenges of police interrogation, focusing on the time-consuming manual procedures and administrative burdens that motivate the need for a digital solution. Section 5 introduces the INTU-AI system, presenting its overall architecture, core functionalities, and the user interface designed to support investigators. Section 6 overviews the artificial intelligence models implemented, covering the specific approaches used for automatic speech recognition, text-based analysis, and multimodal emotion recognition. Section 7 presents the results of a 12-question questionnaire answered by users of the INTU-AI application. Section 8 discusses the results of the system's validation, including transcription accuracy and the practical implications of its use. Finally, Section 9 concludes the paper by summarizing key contributions and outlining possible directions for future work.

## 2. Related Work

This section examines scientific work related to assessing human emotions and cognitive conditions for deception detection. This area has been a subjective exercise heavily reliant on the experience and intuition of human investigators [4]. Research indicates that a human observer's accuracy in identifying deception is only slightly better than chance [8]. The recent advances in AI have triggered interest in utilizing AI as a means to improve investigative methods. Indeed, AI models are capable of identifying indicators of deception such as speaking patterns [4].

In their study, Kleinberg and Verschuere [9] noted that when machine learning was used independently, it was able to determine truth-tellers and liars with an accuracy of 69%. On the other hand, when humans were involved through hybrid-overrule decisions, the accuracy dropped to 50%, equivalent to chance level. An explanation is that individuals rarely face deception and are more inclined to trust someone than to doubt their honesty. This truth bias results in generally greater precision in identifying truths as opposed to detecting lies.

The Autonomous Interrogator and Deception Identifier (AIDI) [10], an autonomous agent powered by Large Language Models (LLMs) such as GPT-3.5 and GPT-4, is designed

to conduct interrogations and assess the truthfulness of statements. AIDI achieves a remarkable accuracy of 77.33% in text-based deception detection, significantly outperforming both human performance (56%) and other LLMs (33% for ChatGPT-4) on this task. Despite offering enhanced objectivity and efficiency, autonomous agents like AIDI lack the empathy and intuition necessary to interpret subtle emotional cues and non-verbal communication, factors that are essential for building rapport and eliciting accurate information. This underscores a significant limitation in comparison to human questioners, who, despite their emotional intelligence, can suffer from stress-related disadvantages [11].

The viability of deception and emotion detection models remains a subject of intense debate, particularly regarding their legal admissibility and accuracy. Brennen and Magnussen [12] highlight these concerns, arguing that the widespread belief in identifying liars through non-verbal cues, such as avoiding eye contact, is a culturally ingrained myth lacking empirical support. Scientific evidence indicates that such cues are weak and unreliable. Although verbal cue analysis achieves approximately 70% accuracy in distinguishing lies from truthful statements, this falls significantly short of the legal threshold of “beyond reasonable doubt”, which is approximately 90–95%. Similarly, physiological methods like polygraphs and neuroscientific techniques such as functional magnetic resonance imaging (fMRI) are plagued by high error rates and substantial theoretical and practical limitations [13]. Additionally, the authors highlight that the strategic use of evidence, gradually introducing independent information to reveal inconsistencies, has shown empirical effectiveness in detecting deception. While machine learning offers promising improvements over human performance, especially in identifying subtle patterns, its use in forensic settings requires caution due to current error rates and limited real-world readiness.

In another study [14], Brennen and Magnussen argued that traditional verbal and physiological methods do not meet legal standards due to high error rates and limited reliability. Although the strategic use of evidence stands out as the most promising human-led approach, it lacks scalability for broader application. Machine learning offers significant potential, with some studies reporting accuracy above 90%, but faces critical limitations, including dependence on simulated data, fairness issues, and susceptibility to manipulation. Ultimately, the most reliable method remains identifying inconsistencies between a subject’s statements and independent evidence.

The notion of complexity is a common thread in most research in the field, as exemplified by the work of Alex Constâncio [15]. The author reports a human deception detection accuracy of approximately 54% and challenges the widespread belief in nonverbal cues, such as avoiding eye contact, as scientifically unreliable. The study highlights a crucial insight: although deception is theoretically associated with emotional states like guilt, fear, or duping delight, most machine learning research has concentrated on behavioral manifestations of these emotions rather than modeling the emotions directly, revealing a notable research gap. Progress in the field is further hindered by the scarcity of real, labeled interrogation data, with over 70% of studies relying on simulated scenarios. Additionally, the opaque nature of black-box models and the predominance of English-language datasets (75%) emphasize the need for more robust research using real-world, multimodal, and multilingual data to ensure the development of practical and trustworthy ML-based deception detection systems.

In this context, video-based deception detection has emerged as a promising field, leveraging multimodal analysis across visual, audio, textual, and physiological inputs. Several studies have highlighted the potential of deep learning models, which consistently outperform traditional methods, particularly when employing multimodal fusion strategies, as demonstrated in the literature review of Rahayu et al. [16]. Despite these advancements, significant challenges remain, including the scarcity of large, diverse, and culturally rep-

representative datasets, class imbalances, difficulties in establishing reliable ground truth, and the complexity of modeling temporal dynamics in real-world interactions. Moreover, ethical concerns, ranging from privacy and bias to interoperability and the psychological impact of deceptive research designs, further hinder practical deployment. Addressing these issues requires a concerted research effort focused on building ecologically valid datasets, improving the model explainability, and ensuring fairness across demographic groups to support the development of trustworthy and socially responsible deception detection systems.

Patel et al. [17] present a state-of-the-art multimodal deception detection system that integrates visual, vocal, and textual cues using deep learning and early fusion strategies. Although it does not focus explicitly on emotions as output categories, it analyzes behavioral and physiological indicators, such as microexpressions, vocal stress, and linguistic patterns, that are often linked to emotional states or efforts to conceal them. The system employs real-time pipelines for facial tracking, speech analysis, and text processing, combining these modalities in a unified architecture using Bidirectional Long Short-Term Memory (BiLSTM) and deep audio-visual models. Trained on datasets like Dolos and PolitiFact, the model achieved strong performance, namely an 85.12% accuracy and an 83.98% F1-score.

### 3. The Need for Digitalization of Police Interrogation

The process of criminal police interrogation can be broadly categorized into three distinct phases, as follows: (1) Information gathering; (2) Interrogation; (3) Report writing. Initially, the process focuses on acquiring details regarding the individual who will be questioned. This initial stage is the least complex aspect of the procedure, primarily involving the collection of identifying information about the individual or any relevant case specifics crucial to the investigation, a task mainly handled by humans. The subsequent phases include the interrogation itself and, finally, the completion of reports related to the interrogation for procedural purposes. These latter functions are fully performed by human personnel. This is where the INTU-AI system intervenes, aiming to digitize the administrative process and assist the interrogator throughout the interrogation.

The interrogation phase is always conducted by a specialized investigator, with sessions typically recorded on video; however, interrogations do not always take place in settings specifically designed for this purpose. They are often conducted in police units, stations, or, in certain cases, in operational theaters near conflict zones, where conditions can vary widely and may include improvised environments like tents or makeshift facilities. Indeed, depending on the place where the interrogation takes place, lighting may influence the image quality of the video, as well as the audio caption and the available recording equipment.

Besides the interrogation itself, the investigator is frequently required to analyze a substantial amount of supplementary material, such as wiretaps or recorded telephone conversations, all of which constitute critical evidence that must be considered during the investigation process. This often occurs before the interview(s).

Once the interrogation concludes, investigators face considerable administrative burdens, largely stemming from the need to document a vast amount of evidence. This includes preparing various legal documents required for judicial proceedings. These documents must not only consolidate factual elements, such as transcripts of wiretaps, phone calls, and interrogations, as well as summaries of investigative actions, but also incorporate identification details of both the interviewee and the interrogator. Currently, this entire process is performed manually and relies entirely on human effort. This procedure, both in terms of forensic analysis and administrative processes, has increasingly led investigators to spend many hours working on interrogation-related case files. A considerable amount

of time is consumed by tasks such as transcription and summarization of content for the official interrogation reports.

It is in this context that INTU-AI emerges as a tool designed to alleviate the workload pressure on investigators by automating and streamlining these administrative time-consuming tasks, resorting to AI. Simultaneously, it provides a decision-support system during the interrogation process by identifying and signaling potentially inappropriate or inconsistent elements, linked to emotions, in real time, again based on AI.

#### 4. Emotion Analysis in Interrogation Contexts

During the interrogation process, the lack of linearity in the responses or statements provided by the interviewees often leads to the involvement of specialists to analyze certain behavioral characteristics, with the aim of analyzing behavioral and emotional patterns of the interviewee. These characteristics can range from body posture to arm and hand positioning, as well as the emotional responses of the interviewee, visible through facial expressions. As stated earlier, emotions have long been a subject of study by prominent scholars [18–20], who have studied emotional indicators as relevant behavioral signals [21–23].

The study of emotions comprises a vast area of research. In interrogation scenarios, the emphasis is primarily on observing facial expressions, the manner of statement delivery, and the actual content of statements [24]. According to DePaulo [25], behavioral signals, including facial expressions, prosody, and verbal content, provide relevant information for emotion analysis. Likewise, Aldert Vrij [26] highlights the practical use of emotional and behavioral analysis in interviews and interrogations, considering visual, vocal, and verbal cues as essential information sources. These dimensions constitute the three main vectors for emotion analysis in the present work: FER, SER, and TEA. Analyzing these aspects is of critical importance, particularly considering that in many cases, certain individuals can regulate their emotional responses during an interrogation through techniques, for example, expressive suppression [27]. It is crucial to recognize that while suppressing emotions might lessen external expressiveness, it does not completely eradicate the emotions themselves, nor does it necessarily diminish the internal emotional experience or physiological responses.

#### 5. The INTU-AI Interrogation System

As stated earlier, INTU-AI is a solution designed to digitize the entire interrogation process while simultaneously providing visual support to interrogators in capturing the emotional states of interviewees during interrogations. It also serves as a decision-support tool. The system aims to keep the human element actively involved throughout the criminal interrogation process, while freeing investigators to focus on the most critical aspects of the procedure.

INTU-AI is a standalone software designed for Windows 11. In terms of hardware, it requires an NVidia GPU with at least 4 GiB of GDRAM, a mid-level x86-64 CPU, and 16 GiB of RAM for optimal performance. In addition to standard peripherals such as a keyboard and mouse, the system also requires a USB video camera and a USB microphone, either as separate devices or integrated into the camera. Important to note, the INTU-AI system functions offline, needing no Internet connection. The list of the libraries employed in the implementation of the application is presented in Table 1.

INTU-AI is specifically designed for single-user operation through a centralized interface. It processes interrogation videos as input, generating analytical reports and emotion-tagged video outputs to assist the interrogator's evaluations. The main inputs, outputs, main functions of the INTU-AI system are detailed next.

**Table 1.** Libraries used by category (the libraries without versions are part of the Python Standard Library).

Artificial Intelligence (AI)	Machine Learning (ML)
whisper (20250625)	tensorflow, tensorflow.keras
transformers (4.46.2)	torch (2.5.1+cu121)
speech_recognition (3.11.0)	joblib (1.4.2)
pyannote.audio (3.3.2)	
spacy (3.8.4)	
pycallgraph2 (1.1.3)	
Data Science	
collections	pandas (2.2.3)
datetime	pdf2docx
docx (1.1.2)	pdfplumber (0.11.4)
fitz (PyMuPDF) (1.24.13)	PyPDF2 (3.0.1)
fpdf (1.7.2)	re (2.2.1)
geocoder (1.38.1)	reportlab (4.2.5)
glob	shutil
librosa(0.10.2.post1)	string
librosa.display	subprocess
logging (0.5.1.2)	sys
matplotlib (3.9.2)	tabula (2.10.0)
matplotlib.pyplot	time
numpy (1.26.4)	webbrowser
os	

### 5.1. INTU-AI Inputs

The user is required to submit two essential inputs to the system: (i) a PDF document bearing the individual's identity information, which could be either a citizenship card or a military registration form, and (ii) one of the following three media options: a previously recorded interrogation video, a pre-recorded audio file, or a real-time live stream from the local camera and microphone being used during the interrogation session. This allows the system to support both real-time interrogation scenarios and post-interrogation analysis. The video input serves two purposes: (i) Sentiment analysis to provide the interrogator with clues to detect possible deception; (ii) Transcription of the audio to be included in the final report, as well as text sentiment analysis. Note that the application can also be used to transcribe any audio file, for example, of a surveillance recording or of a phone wiretapping.

### 5.2. INTU-AI Outputs

All outputs are delivered in a local folder on the user's device. The system provides five main output items: (i) A detailed report indicating the top emotion detected in each ten-second time windows, with all segments labeled according to the three emotion vectors: SER, FER and TEA; (ii) A fully labeled video of the interrogation, where the predominant emotion from the three vectors is annotated throughout. This enables the investigator to visually review the entire interrogation and pinpoint specific emotional states at any given moment; (iii) A final report generated automatically to comply with administrative procedures, including process identification and the identification details of both the interrogator and the interrogated person, all extracted from the submitted identification documents; (iv) A complete transcription of the interrogation audio, along with a summarization of its content; (v) The results of the emotion analysis for each of the three emotion vectors are provided in both a PDF report and on a video file. The three sections of a PDF report example are illustrated in Figures 1–3, while a frame of the MP4 video file is shown in Figure 4. For INTU-AI, and using the PJM as a case study, the automatically produced report aligns with the agency's currently used official documentation templates.

To summarize, by supplying either an audio/video file or a live stream and an ID document of the person in question, investigators allow the system to automatically finalize all required paperwork, customized according to the official PJM templates. Furthermore, the system produces a functional report featuring a transcript of the input audio/video or stream, a synopsis of its content, and generates an output video/audio marked with the predominant emotions identified across FER, SER, and TEA.



**MINISTÉRIO DA DEFESA NACIONAL  
FORÇAS ARMADAS  
ESTADO-MAIOR DAS FORÇAS ARMADAS  
RELATÓRIO DE ENTREVISTA - PJM**

**DADOS DO INTERROGADO**

Posto: NIM: Nome: José António da Silva Pinto Garcia  
Data de Nascimento: Information intentionally omitted  
Cartão de Cidadão: Information intentionally omitted Validade: DD-MM-YYYY  
NIF: Information intentionally omitted Validade

**DADOS DO INTERROGADOR**

Interrogador: None  
Data do Interrogatório: 14/02/2025, Horas: 21:03  
Local: PJM Lisboa , Cidade: None

Aos 14/02/2025, no PJM Lisboa em None compareceu perante mim, None, o José António da Silva Pinto Garcia portador do Cartão de Cidadão n.º (Information intentionally omitted Validade), válido até DD-MM-YYYY. Declarou ser filho de (Information intentionally omitted Validade) e de (Information intentionally omitted Validade), ter Information intentionally omitted anos de idade, residir na (Information intentionally omitted Validade).--/

**RESUMO DA ENTREVISTA**

A juíza americana Mindy Glaser ficou famosa quando reconheceu um colega de escola durante

**Figure 1.** First section of the final report (in Portuguese) containing the identification of the interviewee and interviewer, general information about the interrogation, and a brief summary.



**MINISTÉRIO DA DEFESA NACIONAL  
EXÉRCITO PORTUGUÊS  
ESTADO-MAIOR GENERAL DAS FORÇAS ARMADAS  
RELATÓRIO DE ENTREVISTA - PJM  
ANEXO A - TRANSCRIÇÃO DO INTERROGATÓRIO**

A juíza americana Mindy Glaser ficou famosa quando reconheceu um colega de escola durante um julgamento. As imagens do emocionante reencontro rodaram o mundo. Agora ela teve um novo reencontro no tribunal onde trabalha, na Flórida. Alan Gleam foi preso acusado de cometer fraude. Semanas antes, ele estava em um cruzeiro. A juíza também. Em junho, Glaser havia reconhecido Arthur Boff, que havia sido preso ao fugir da polícia após um furto. Oh meu Deus! Eu estou muito desculpada por ter visto você lá. Eu sempre me perguntava o que aconteceu com você, senhor. Ela e o Hell costumavam jogar futebol juntos no ensino médio. Esse é o mais bonito garoto da escola. Oh meu Deus! Ele era o melhor garoto da escola. Eu costumava jogar futebol com ele e com todos os garotos. E veja o que aconteceu. Eu estou muito desculpada. Oh meu Deus!

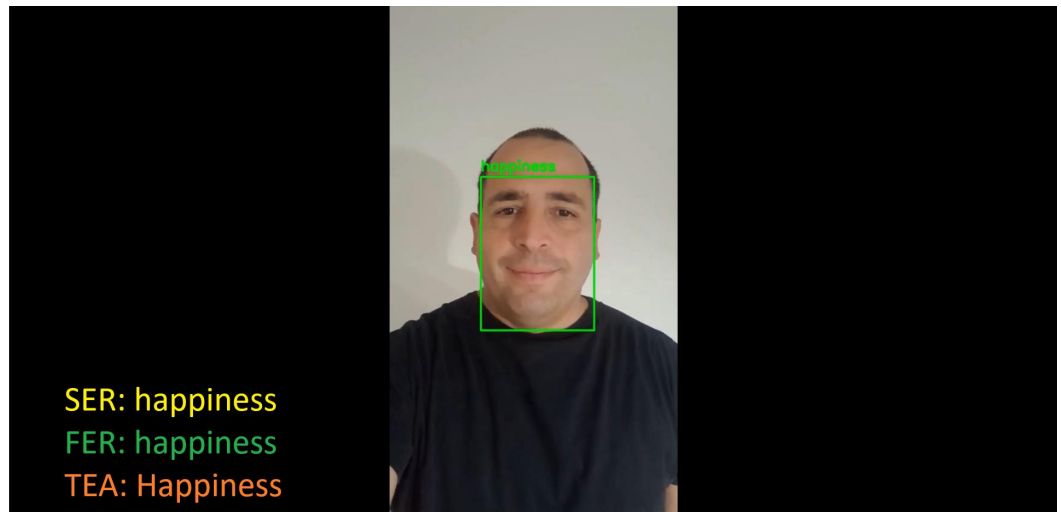
**Figure 2.** Second section of the final report (in Portuguese) providing the transcription of the interrogation.



**MINISTÉRIO DA DEFESA NACIONAL  
EXÉRCITO PORTUGUÊS  
ESTADO-MAIOR GENERAL DAS FORÇAS ARMADAS  
RELATÓRIO DE ENTREVISTA - PJM  
ANEXOS B, C e D - SUMMARY AER, FER e Sentiment Analysis**

Tempo (SER)	SER	FER	Sentiment Analysis
(0s-10s)	surprise	happiness	Happiness
(10s-20s)	surprise	neutral	Happiness
(20s-30s)	surprise	happiness	Happiness
(30s-40s)	surprise	neutral	Happiness
(40s-50s)	surprise	neutral	Anger
(50s-60s)	surprise	neutral	Anger
(60s-70s)	surprise	neutral	Anger
(70s-80s)	surprise	neutral	Sadness
(80s-90s)	surprise	neutral	Happiness

**Figure 3.** Third section of the final report presenting a summary table of the emotions extracted from the three analyzed vectors.



**Figure 4.** Excerpt from a video processed by the INTU-AI program, showing in the lower-left corner the predominant emotions (FER, SER, and TEA) analyzed over a 10-s interval.

**5.3. INTU-AI Main Functionalities**

With an interactive interface, the entire INTU-AI system, including both backend and the TKinter-based graphical interface (GUI) front-end, was developed in Python (version 3.13.2) to meet the specified technical requirements and operational needs. The main interface layout and its functionalities are illustrated in Figure 5.

A new interrogation process begins with the completion of administrative data, specifically the personal identification of the interviewee. This includes full name, date and place of birth, identification photo, parentage, and address. To ease this process, INTU-AI can automatically extract and populate this information from a scan of the provided identification document, such as the Portuguese Citizen Card (Figure 6) or Military form, streamlining the process and minimizing manual input. The user can review all extracted data in the upper left section of the interface as shown in Figure 7.

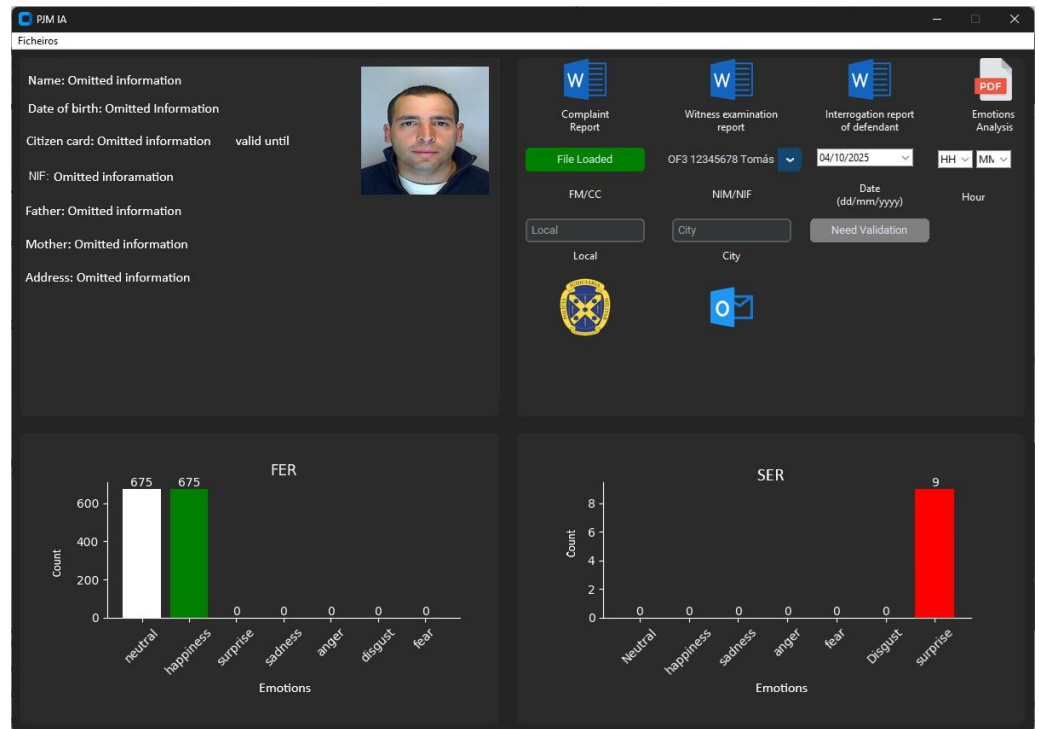
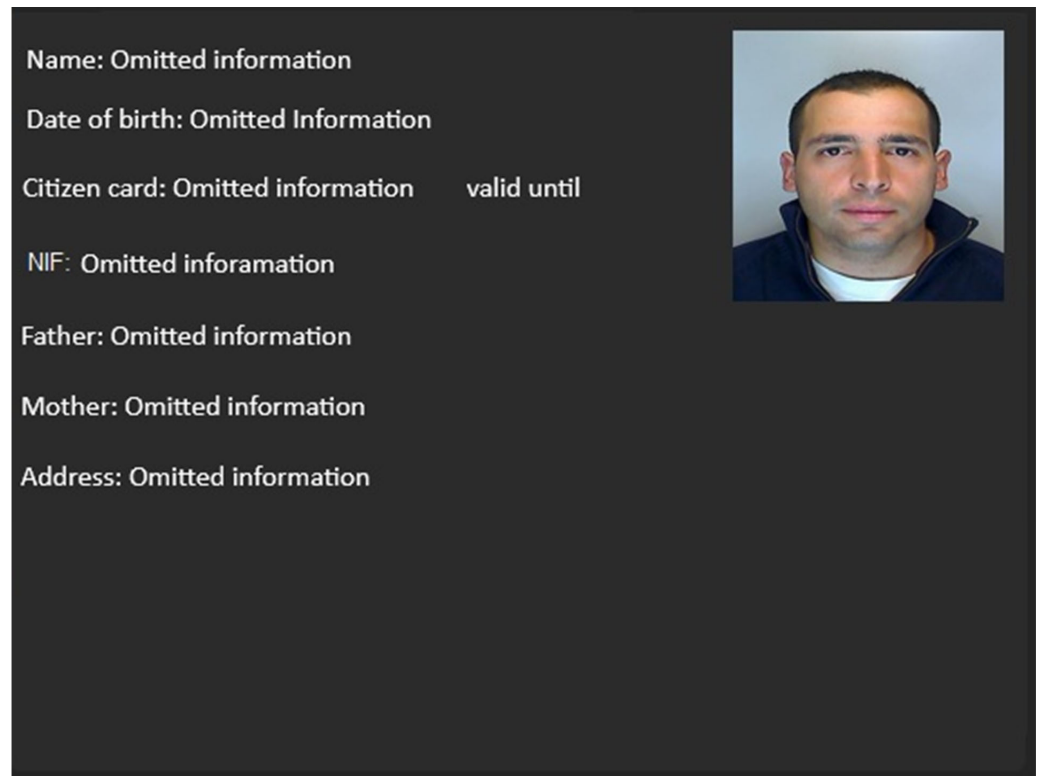


Figure 5. INTU-AI Main Menu Layout interface.



Figure 6. Citizen Card PDF [28].



**Figure 7.** Personal information extracted from input identification documents.

During an interview, the interface allows the interrogator to visually track the overall FER and SER emotions throughout the interrogation via two graphs located at the lower right and left corners, respectively, as shown in Figures 8 and 9. The GUI also provides an interactive panel, to allow users to access various available reports, including the final report of the entire process created by the application. The drop-down menu in the upper left corner enables the user to input data, whether audio, video, or streaming. Regarding processing speed for the mode where the researcher provides a video or audio input, the complete workflow, from inputting the subject's ID to receiving the annotated video and associated reports, typically takes approximately 30 min for a 5-min recording, that is a 6-to-1 ratio. These performance metrics were gathered using a standard laptop with a 12th Gen Intel® Core™ i5-12450H2.00GHz CPU, 16.0 GiB RAM, and an NVidia GeForce RTX 4050 GPU.

At present, the system operates in a post-hoc manner, capturing and processing recorded video rather than performing real-time analysis. In other words, the prototype functions as if pre-recorded interrogation sessions were provided as input. Future work will focus on optimizing model efficiency and exploring lightweight architectures to bridge the gap between offline analysis and real-time applicability.

Regarding privacy by design, the program operates in a completely offline environment. In practice, it is deployed on a machine provided by the Armed Forces, where initial authentication is performed by the user through multi-factor authentication. Therefore, to use this program, a user must have authorized access to the PJM's machines and complete the multi-factor authentication process. File retention policies are defined by the PJM and are aligned with the organization's institutional framework, as well as with the Portuguese Constitution and the GDPR. Likewise, the capture of images is clearly regulated under Portuguese law, specifically in the Code of Criminal Procedure [29].

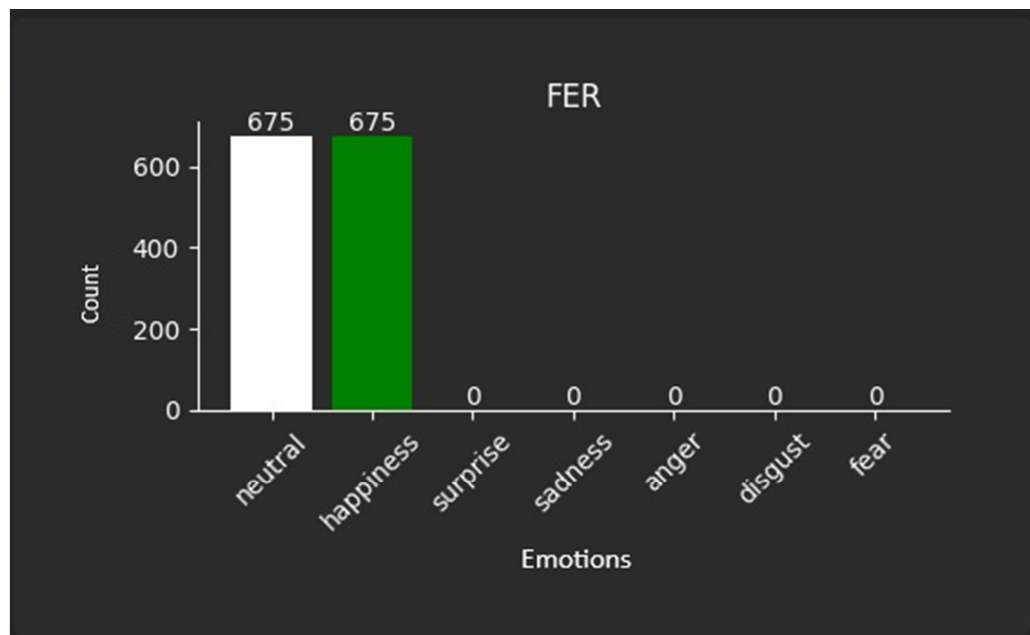


Figure 8. Face Emotion Recognition—Emotion count plot.

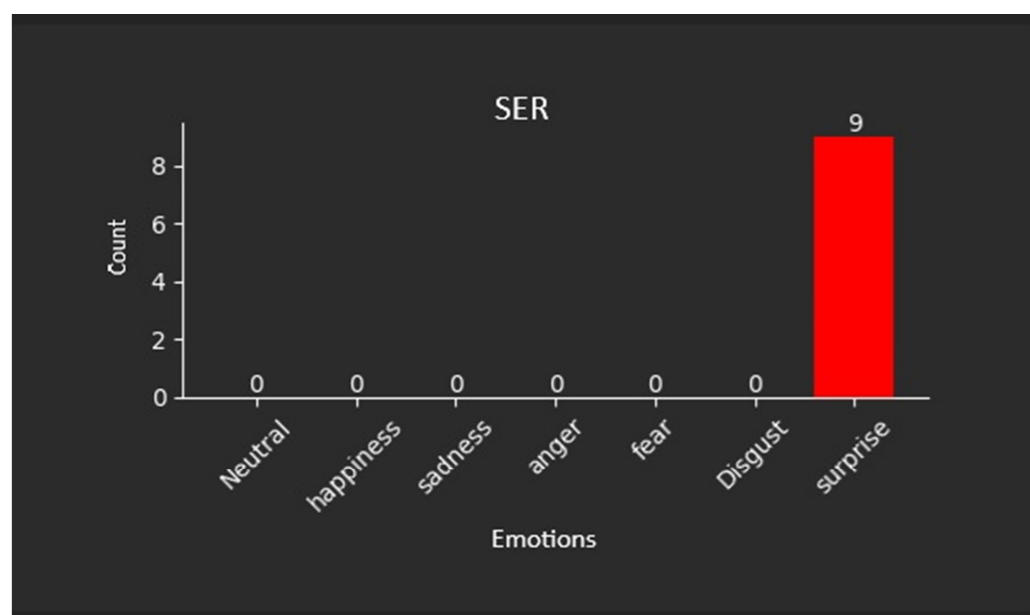


Figure 9. Speech Emotion Recognition—Emotion count plot.

## 6. Artificial Intelligence Usage in INTU-AI

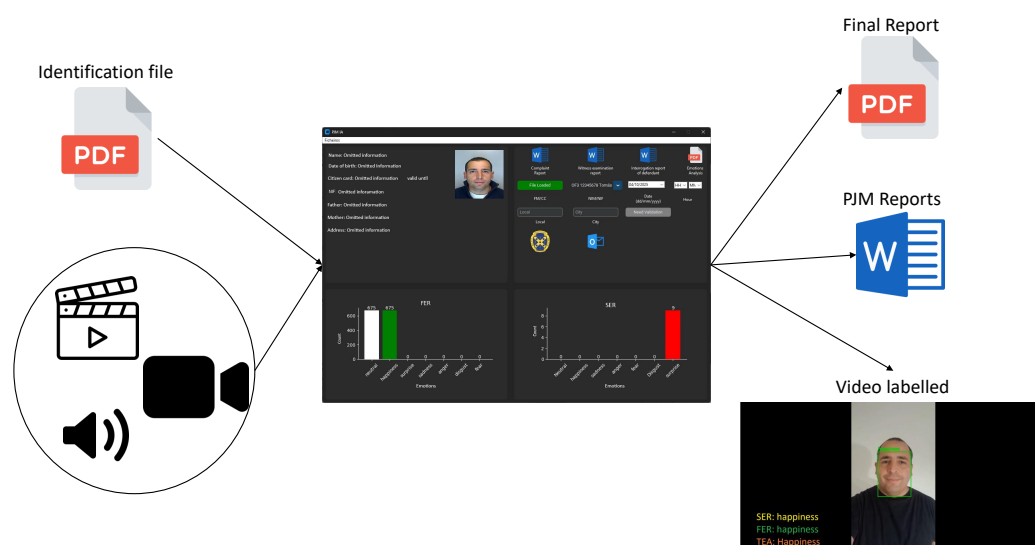
To transform spoken audio into text, that is, perform automatic speech recognition (ASR), which is necessary for interviews, transcriptions, and related natural language processing tasks, the INTU-AI software operates using a step-by-step procedure. Specifically, it resorts to the open-source tool Whisper [30,31] for transcribing audio into text. This transcription is multifunctional, as it is included in the appendix of the official document, acts as input for the summarization model, and forms the textual groundwork for classifying emotions.

### 6.1. Natural Language Processing

For natural language processing (NLP), one of the approaches implemented was text summarization using the open-source model `flan-t5-portuguese-small-Summarization` [32].

Given that the system is intended for use by two Portuguese law enforcement agencies, the selection of AI models has consistently prioritized compatibility with the Portuguese language. In addition, emotion classification from text was also employed, aiming to categorize emotional content according to Paul Ekman’s six basic emotions, that is, (i) anger, (ii) disgust, (iii) fear, (iv) happiness, (v) sadness, and (vi) surprise [19]. To achieve this, various methodologies were explored, including traditional machine learning algorithms, deep learning architectures, and transformer-based models.

To extract data from the PDF version of user-provided identification documents—name, ID number, etc.—we applied regular expressions, rule-based methods, as well as the Python’s PyMuPDF library [33]. All this structured information, extracted through ASR, NLP techniques, and document parsing, is used to populate the system interface and to automatically generate the official interrogation reports, as illustrated in Figure 10.



**Figure 10.** Information extraction flux on INTU-AI.

### 6.2. Multimodal Emotion Recognition

To perform multimodal emotion recognition, we integrated specialized models for each modality—facial expressions, vocal tone, and textual content—within a unified pipeline. For facial emotion recognition, we employed a combination of two open-source models: a facial detection (FD) model [34] and an FER model [35]. These models work jointly to capture the visual perception of the subject’s emotional state in real time. For this purpose, the video is segmented into 10-s chunks with the FER classification applied to each one of these chunks. For SER, we used the open-source model wav2vec2-lg-xlsr-en-speech-emotion-recognition [36], which is a fine-tuned version of a pre-existing architecture [37], specifically designed for detecting emotional patterns in voice tone. Most SER models are trained on English corpora, which creates language and domain gaps, particularly in prosody, phonetics, and cultural emotion expression, when applied to Portuguese, reducing generalization and accuracy.

In the textual modality, the lack of emotion analysis datasets in Portuguese required the adaptation of an existing resource, specifically, the AffectAlchemy dataset [38] to the Portuguese language. For this purpose, in an initial phase, automated translation tools were employed, namely DeepL [39], and deep-translator [40] to transcribe and linguistically adapt the dataset. Upon concluding this pre-processing phase, we proceeded to train our emotion classification model.

To ensure consistency between datasets, the AffectAlchemy corpus (originally 20,075 observations with “Text” and “Emotion”) was harmonized with the TEA framework. While TEA follows Ekman’s six basic emotions, AffectAlchemy is based on Plutchik’s model; therefore, we applied a systematic mapping between the two schemes, complemented by the inclusion of a neutral category. After filtering the dataset to retain only the target emotions relevant to our multimodal pipeline (Sadness, Happiness, Disgust, Anger, Fear, Surprise, and Neutral), the final corpus was reduced to 8988 observations, ensuring compatibility and interpretability across modalities.

A comprehensive preprocessing pipeline was applied to ensure data quality and consistency. The main steps included text normalization (lowercasing, punctuation and stopword removal, emoji conversion), spelling and grammar correction (via LanguageTool and heuristic rules), tokenization and lemmatization, and hashtag decomposition using spaCy and wordninja. Named Entity Recognition (NER) was also performed, with optional anonymization to preserve privacy. Additionally, sentiment analysis was conducted using TextBlob and VADER to obtain polarity and subjectivity scores.

The dataset was translated into Portuguese using Google Translate, followed by manual correction of 254 sentences that were not automatically processed. After post-processing, the Portuguese dataset comprised 8,988 sentences with a total of 57,619 words. Sentiment analysis revealed a slightly positive profile (polarity approx. 0.056) and a moderately subjective tone (subjectivity approx. 0.559).

NER analysis identified 18 entity types, with the four most frequent—CARDINAL, PERSON, DATE, and ORG—accounting for 76% of all mentions. This indicates a predominance of numerical, temporal, and organizational references.

Subsequently, to determine the most appropriate algorithm for emotion classification, we applied a range of traditional machine learning algorithms, including Support Vector Machines (SVC), Naive Bayes, Random Forest, Decision Tree, XGBoost, K-Nearest Neighbors (KNN), and Logistic Regression. We also explored more advanced approaches, implementing two deep learning models, a Convolutional Neural Network (CNN) and a Recurrent Neural Network (RNN), both leveraging pre-trained FastText embeddings [41]. Finally, to further enhance performance, we fine-tuned state-of-the-art transformer-based models, namely DistilBERT and RoBERTa [42], using the AffectAlchemy dataset.

The performance of all assessed models, evaluated with the F1-score metric on the validation set, is summarized in Table 2. Among them, the transformer-based model RoBERTa stood out by achieving the highest F1-score. This metric is especially pertinent due to the moderate class imbalance in the dataset. For instance, the emotion “joy” occurs nearly thrice as often as “trust”. Consequently, the F1-score provides a more comprehensive and equitable evaluation of the model’s performance across all emotional categories.

**Table 2.** Validation metrics from models used on Portuguese translated AffectAlchemy dataset.

Model/Metrics	F1
Reg. Log.	0.6648
NB	0.6530
XGBoost	0.6196
DT	0.5075
RF	0.5878
KNN	0.5094
SVC	0.6361
CNN + FastText	0.5687
RNN + FastText	0.5876
RoBERTa	0.6715
DistilBERT	0.6270

The results presented above correspond to the performance of various models applied to the dataset made available by the authors [43], which contains approximately 10,000 observations. When replicating the experiments using the same dataset, we obtained slightly different results from those originally reported. These are presented in Table 3. For comparison, we also include the performance of a RoBERTa-based model trained on the same dataset in English.

**Table 3.** Performance of models on the Affect Alchemy dataset.

Model	Accuracy	F1-Score
Baseline (replication)	0.6618	0.6655
RoBERTa (English)	0.8582	0.8572

### 6.3. Audio Transcription

From the human resources point of view, one of the key benefits provided by AI usage in INTU-AI is the functionality of transcribing the interrogation, freeing personnel power from the cumbersome task of manual transcription of the audio, as human operators only need to validate the transcription, correcting transcription errors. To evaluate transcription quality, we considered the Word Error Rate (WER) of the Whisper Large model. According to OpenAI's original benchmarks, Whisper achieves a WER of 4.1% on English audio (LibriSpeech test-clean), while for Portuguese and other non-English languages, the WER typically ranges between 6% and 10%, depending on audio quality, speaker accent, and background noise [31].

Due to the lack of actual interrogation data, we executed an internal evaluation employing synthetic material. We created 34 fake interrogation transcripts using ChatGPT and transformed them into video format with the open-source tool InVideo AI [44]. These videos [45] were then analyzed by the INTU-AI audio-to-text transcription system to assess its efficiency. Due to the synthetic content from ChatGPT containing a large amount of additional information, including environmental details and insights into the psychological or emotional conditions of both the interrogator and the interviewee, we implemented a preprocessing stage to filter the material. Particularly, since the initial texts were in .docx format, we executed a preprocessing stage [46] to detach the interrogation dialogue from the contextual narrative. After extracting the interrogative sentences, the WER of the audio-to-text transcription was calculated automatically using the JiWER Python library [47].

To evaluate the transcription system in a more authentic, non-LLM-generated setting, we implemented the same method on 10 episodes of the Portuguese legal drama *A Sentença* [48]. Since these episodes generally last over an hour, we specifically chose portions containing direct dialogue between the judge and participants such as witnesses, defendants, or offenders. In this instance, the WER was assessed manually, owing to the lack of an official transcript. A human evaluator made comparisons between the transcription output and the spoken dialogue. The mean WER was found to be 13.22%, showing a slight improvement over the synthetic data test, possibly due to differences in background noise, voice clarity, and intonation found in the AI-generated video.

### 6.4. Text Summarization

Beyond transcription, the AI system also provides automatic summarization of the transcribed audio. Another key benefit from the human workload point of view, although not AI-dependent, is the handling and creation of the formal administrative documentation required for the process, substantially reducing the human workload associated with post-interrogation reporting.

To evaluate the performance of the used summarization model [32], we adopted an experimental methodology based on comparison with the summarization capabilities of two general-purpose models: OpenAI's ChatGPT [49] and DeepSeek [50]. Both were accessed through prompt-based instructions. This procedure resulted in three summarized versions for each of the 34 simulated interviews. To assess the quality of the generated summaries, we applied the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) [51] standard evaluation metrics, namely, ROUGE-1, ROUGE-2, and ROUGE-L. The ROUGE-n metric measures the overlap of n-grams, that is, the contiguous sequences of n items, such as words or characters, between the system-generated and reference summaries. ROUGE-1 refers to the overlap of unigrams (single words), ROUGE-2 refers to the overlap of bigrams (pairs of words), and so on. ROUGE-L measures the longest common sequence (LCS) between the system-generated and reference summaries, with LCS being the longest sequence of words that appears in both summaries in the same order, but not necessarily contiguously. The ROUGE results from our experiments are shown in Table 4.

**Table 4.** Consolidated validation results across INTU-AI modules: transcription, summarization, and user questionnaire. *S* is the number of substitutions, *D* deletions, and *I* insertions.

Validation Domain	Test/Model	Metric	Result
Transcription	Synthetic interrogations (34)	WER	15.93% <sup>1</sup>
	Legal drama series A Sentença (10)		13.22% <sup>2</sup>
Summarization	ChatGPT	ROUGE-1/ROUGE-2/ROUGE-L	0.1619/0.0748/0.1120
	DeepSeek		0.1116/0.0444/0.0800
	LAN T5		0.0807/0.0666/0.0768
Questionnaire (Likert 1–5)	Usability (Items 1–3)	Agree/Strongly Agree	78%
	Functionality (Items 4–6)		74%
	Practical Utility (Items 7–9)		81%
	User Experience (Items 10–11)		70%

<sup>1</sup> WER 95% CI: 15.58–16.28%, S/D/I: 1.37/1.03/1.09. <sup>2</sup> WER 95% CI: 12.85–13.59%, S/D/I: 0.49/1.84/0.72.

Our experimental evaluation using ROUGE metrics reveals substantial performance differences among the three text summarization models under study. ChatGPT outperforms the others across all metrics, achieving ROUGE-1, ROUGE-2, and ROUGE-L F-scores. These results highlight its effectiveness in preserving key content units, with a 45% advantage over DeepSeek's ROUGE-1 score. The 68% increase in ROUGE-2 performance over DeepSeek and LAN T5 suggests a higher capacity for capturing phrasal coherence and local syntactic patterns. ROUGE-L results also favor ChatGPT, followed by DeepSeek and, at last, LAN T5, indicating superiority in structural alignment and content coverage.

Notably, the overall low absolute ROUGE-2 values across all models align with findings in the literature, which emphasize the challenges of maintaining exact phrase-level overlap in abstractive summarization. Nevertheless, considering its open-source nature, the performance of the selected model [32] can be regarded as sufficiently acceptable for the task of transcribing and summarizing the simulated interviews.

An important contribution of AI usage is sentiment analysis, which occurs in the FER, SER, and TEA modalities. This multiple approach to sentiment analysis helps to detect instances where subjects attempt to conceal their true feelings. By providing distinct emotional vectors, the system provides cross-verification among modalities, thereby increasing the likelihood of detecting attempts at emotional masking. When used in streaming mode, INTU-AI also serves as a tool to dynamically help to explore the emotional states of interviewees, potentially guiding investigators toward new lines of inquiry during the interrogation.

## 7. Validation and User Feedback

The INTU-AI software was officially launched as an independent executable (.exe) [52] for the Windows operating system at the end of January 2025. Owing to privacy concerns that made real-world interrogation scenarios unfeasible, system validation was conducted during the PJM inspector training sessions in February and March 2025, involving PJM inspector trainees acting as both interviewees and interviewers. Additional testing is underway, with a focus on system performance, compatibility, security, and data confidentiality, utilizing input from PJM. Presently, the project is in its preliminary post-deployment stage, which involves corrective maintenance and iterative improvements based on user feedback. Next, we present the results of a questionnaire answered by 23 PJM trainees.

### Questionnaire and Results

A 12-question structured questionnaire was developed and administered to assess the system qualitatively [53]. The survey was distributed to 23 instructor officers and investigators from the PJM, with the goal to collect feedback across several dimensions: usability, existing functionalities, practical utility of the program, user experience, and open-ended qualitative responses. The evaluation was conducted individually with anonymous respondents through a shared online link [53].

The assessment instrument employed a 5-point Likert scale [54] shown in Table 5, with items adapted from the validated technology acceptance model [55]. This approach allowed for both quantitative measurement of user perceptions and qualitative analysis of open-ended feedback. Table 6 provides an overview of the dimensions assessed by the questionnaire administered in this study.

**Table 5.** Used Likert Scale.

Scale Value	Interpretation
1	Strongly Disagree
2	Disagree
3	Neutral
4	Agree
5	Strongly Agree

**Table 6.** Questionnaire Structure Overview.

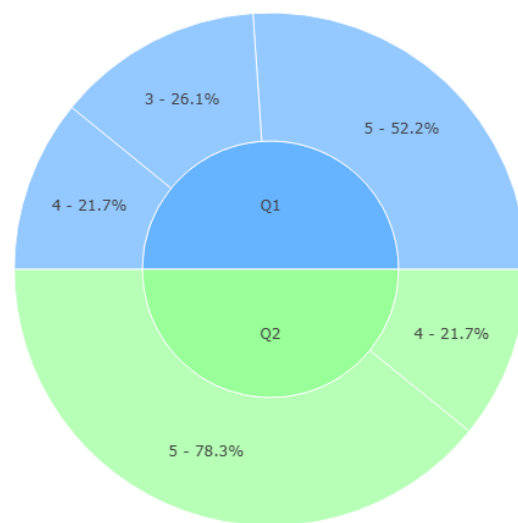
Section Type	Evaluation Dimension
Likert-scale Items (1–9)	Usability Functionality Practical Utility User Experience
Open-ended Questions (10–11)	Suggested Improvements Problem Reporting
Demographic Data (12–13)	Professional Role Experience

The results obtained from the administered questionnaires are presented in Figures 11–14, organized according to the evaluation dimensions described in Table 6. The figures show the percentage of answers that correspond to the respective Likert value. Regarding the evaluation dimensions, specifically Usability, the applied questions were Q1: “The program’s interface is intuitive and easy to navigate?” and Q2: “The processing time

(video/audio analysis) is acceptable for practical use?”. The corresponding results are presented in Figure 11.

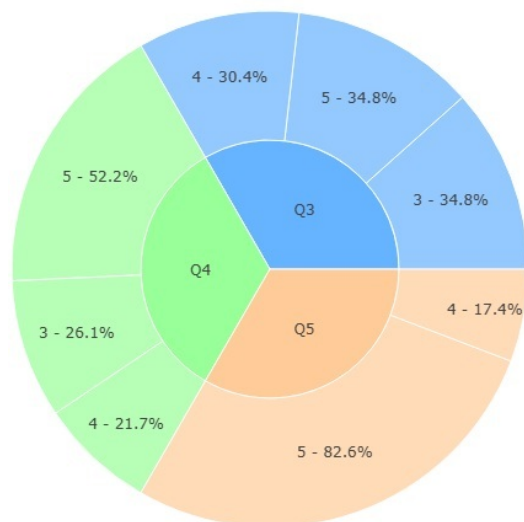
Regarding interface usability (Q1), the responses corresponded to levels 5 and 4 on the Likert scale, indicating a high level of acceptance and perceived intuitiveness of the program. However, 26.1% of respondents selected level 3 (Neither Agree nor Disagree), suggesting that some users may have experienced difficulties. This could potentially be improved through an initial training session or guided introduction, as participants only had access to the program’s user manual, without any direct explanation of how the system operates.

In relation to the next question—Q2: speed of audio/video processing—the obtained results reflect a high level of satisfaction with the processing speed for video and audio analysis.



**Figure 11.** Distribution of answers to questions Q1: “The program’s interface is intuitive and easy to navigate?” and Q2: “The processing time (video/audio analysis) is acceptable for practical use?”.

For the functionality dimension, the related questions were Q3: “Emotion recognition tools (FER/SER/TEA) are accurate?”, Q4: “Automatic transcription of audio accurately reflects the spoken content?”, and Q5: “Automatic report generation (PDF/DOCX) meets operational needs?”. Results for this dimension are presented in Figure 12. These results were generally positive. Regarding Q3, responses were more evenly distributed: approximately 34.8% of participants selected “Strongly Agree” (level 5), and 30.4% chose “Agree” (level 4), totaling 65.2% expressing confidence in the model’s ability to classify the three emotion vectors (FER/SER/TEA). Meanwhile, 34.8% of respondents selected “Neither Agree nor Disagree”, reflecting a degree of uncertainty. This neutrality is further clarified in the open-ended responses, where participants expressed skepticism about the accuracy of the models in the context of criminal investigations, suggesting that future models should be trained on domain-specific datasets to increase trust and applicability.

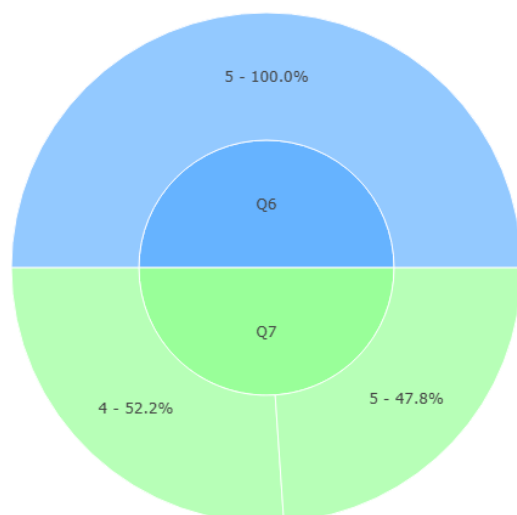


**Figure 12.** Distribution of answers to questions Q3: “Emotion recognition tools (FER/SER/TEA) are accurate?”, Q4: “Automatic transcription of audio accurately reflects the spoken content?” and Q5: “Automatic report generation (PDF/DOCX) meets operational needs?”.

Concerning question Q4—“Does automatic transcription of audio accurately reflect the spoken content?”—a majority of respondents (52.2%) strongly concurred that transcription faithfully mirrors the spoken content. The remaining 47.8% were split between Agree (21.7%) and Neither Agree nor Disagree (26.1%). This also corresponds to the previously evaluated WER and highlights the absence of speaker diarization in the transcript.

Concerning the last inquiry within the functionality domain—Q5—“Does automatic report generation (PDF/DOCX) satisfy operational requirements?”, respondents unanimously agreed that the application adequately produces the required reports and is in line with the operational necessities of the investigative procedures.

The questions addressing the dimension of Practical Utility were Q6: “The program improves efficiency in writing interrogation reports?” and Q7: “The generated emotion graphs are useful for behavioral analysis?”. Results for this dimension are presented in Figure 13.



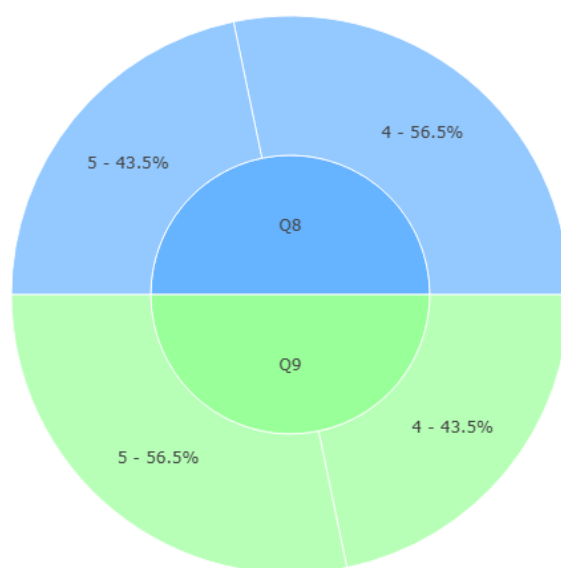
**Figure 13.** Distribution of answers to questions Q6: “The program improves efficiency in writing interrogation reports?” and Q7: “The generated emotion graphs are useful for behavioral analysis?”.

The results indicate a strong acceptance of the INTU-AI system in terms of its Practical Utility. All respondents agreed that the program improves efficiency in writing interrogation reports (Q6), even though some selected a rating of 4 on the Likert scale and identified, in open-ended responses, the need for *speaker diarization* to improve transcription accuracy. Despite these remarks, there was unanimous agreement regarding the program's positive impact on report writing efficiency.

Likewise, for question Q7—“Are the generated emotion graphs beneficial for behavioral analysis?”, the general assessment indicates a positive reception of the tool as an auxiliary resource for investigators, facilitating the tracking of emotional fluctuations during interviews. This graphical depiction of emotional trends was viewed as especially useful for obtaining a comprehensive summary of the subject's behavioral patterns.

This perspective is reinforced by the testimony of the Head of the Criminal Department of the Military Judicial Police, who stated: “...one of the things I end up identifying is that transcription and writing activities are always very time-consuming (...), which, during the evidentiary phase, may affect the course of the investigation”.

Regarding the questions related to the User Experience dimension, namely Q8: “The program reduces human errors compared to traditional methods?” and Q9: “The integration between video, audio and reporting is well implemented?”, the results are presented in Figure 14. The analysis of the User Experience dimension also yielded positive results, with no responses falling below the level of “Agree” for either question. Regarding Q8, participants expressed agreement that the program contributes to a reduction in human error during the interrogation workflow. In open-ended responses, several users emphasized that correcting a machine-generated transcription is considerably easier than performing the transcription manually.



**Figure 14.** Distribution of answers to questions Q8: “The program reduces human errors compared to traditional methods?” and Q9: “The integration between video, audio and reporting is well implemented?”.

In the open-ended questions, specifically Q10: “What additional features would be useful?” and Q11: “Describe any difficulties encountered during use.”, the responses broadly converged on a few key suggestions. As reported earlier, participants emphasized the importance of incorporating speaker diarization to improve the clarity of transcriptions by identifying who is speaking at each moment. Additionally, several respondents proposed extending the emotion recognition capabilities beyond FER, SER, and TEA to include other behavioral cues such as hand positioning or body movement, which could enrich

the multimodal analysis and provide deeper insights into the interviewee’s emotional and cognitive state.

Overall, the qualitative evaluation of the program (Q12) by the respondents was classified as “Very useful”. This perception is reinforced by the statement of the Head of the Criminal Department of the Military Judicial Police, who noted: “(...) this tool [INTU-AI] (...) is an added value as it allows not only the recording and/or audio (...) but also transcribes and fills in the reports, which greatly facilitates the conduct of the investigation (...) in addition to all this, this investigative support tool assists in reading emotions (...) which may allow us to open new lines of investigation.”

### 8. Limitations and Challenges

One limitation observed in the transcription process is that the system does not perform speaker diarization, that is, it does not distinguish between speakers. As a result, the conversation is transcribed as a continuous stream of text without indicating who said what. This requires the interrogator to identify the speaker when post-processing the output text accordingly. We aim to add speaker diarization as future work, exploring libraries that provide this functionality and that run locally, such as PyAnnote and SpeechBrain.

The scarcity of datasets focused on criminal interrogations leads to an inadequacy of models tailored for this context. Furthermore, acquiring such data to train models is difficult because of regulations that limit access to protect data privacy. Consequently, the approach must include a proof of concept utilizing simulated data that mirrors features typical of an interrogation environment. Despite the limitations in using this concept and data for publicly accessible models, it remains possible to develop models intended for private application. Thus, within the INTU-AI framework, the long-term goal is to replace the current FER, SER, and TEA models with new models specifically trained on labeled interrogation datasets, once such data becomes available, ensuring greater contextual accuracy and relevance.

A particular requirement for accurate emotion analysis in text is that the underlying model should ideally be trained on a dataset in Portuguese. However, due to the scarcity of emotion-labeled datasets in this language, the training was conducted on transcriptions originally produced in other languages, like English, Spanish, French, and others, rather than on datasets natively constructed in Portuguese. Hence, for future development, our strategy is divided into three main areas: speaker diarization, development of a dataset geared towards police interrogations and native Portuguese emotion datasets, and improvements to the program itself through body-language indicators and explainable NLP. The prioritization of these strategies is summarized in Table 7.

**Table 7.** Roadmap for future work in INTU-AI application.

Time Frame	Focus Area
Near-term	Speaker diarization
Mid-term	Police interrogations dataset and native Portuguese emotion datasets
Longer-term	Body-language indicators, explainable NLP

### 9. Conclusions

This paper introduced INTU-AI, an AI-driven system designed to address the significant challenges of traditional police interrogation procedures, which are often manual, time-consuming, and heavily reliant on subjective human judgment. By integrating a multimodal approach that analyzes FER, SER, and TEA, INTU-AI provides a comprehensive solution for the digitization and enhancement of the entire interrogation workflow.

The main contributions of this work are twofold. Firstly, the system streamlines and automates cumbersome administrative tasks by providing functionalities for automatic transcription, summarization, and the generation of official, agency-compliant reports. This automation frees investigators from significant bureaucratic burdens, allowing them to concentrate on the critical, analytical aspects of the investigation. Secondly, INTU-AI serves as a decision-support tool. By presenting real-time emotional data through a graphical interface, it aims to assist investigators in identifying emotional inconsistencies and potential attempts at concealing information, thereby contributing to the investigative process. The validation of the system during training sessions with police inspectors and police trainees, although limited, confirmed its practical applicability and potential. Ultimately, INTU-AI helps criminal investigations, augmenting the capabilities of human investigators with objective, data-driven insights while keeping them at the center of the decision-making process.

To ensure practical effectiveness and usability, future work will involve real-world validation, with iterative refinements based on the reported users' feedback. Another goal for the INTU-AI system is to focus on the development of a native Portuguese dataset to train more accurate emotion recognition models, moving beyond the current reliance on translated data. The integration of Large Language Models (LLMs) can enable more advanced functionalities, such as the implementation of an intelligent chatbot to assist investigators. This could be based on self-paced learning approaches, using open-source models such as LLaMA [56], run locally to avoid exposing sensitive data to external services. As stated earlier and requested by most users of the application, we also aim to provide for speaker diarization. Finally, another possible enhancement is to expand the system's analytical capabilities to include other behavioral indicators such as body language and micro-expressions, alongside more advanced NLP techniques for a deeper narrative analysis.

**Author Contributions:** Conceptualization, J.P.G., C.G., R.M. and P.D.; methodology, J.P.G., C.G., R.M. and P.D.; software, J.P.G.; Validation, P.D., R.M. and C.G.; writing—original draft preparation, J.P.G. and P.D.; writing—review and editing, C.G., P.D. and R.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partially supported by the Fundação para a Ciência e a Tecnologia, I.P. (FCT), under project reference UID/50008/2023 IT, and in the context of project UIDB/04524/2020, as well as through the Scientific Employment Stimulus (CEECINS/00051/2018). The authors also gratefully acknowledge the financial support from the European Union via the Portuguese Recovery and Resilience Plan, under the Call: 2021-C05i0101-02-agendas/alianças mobilizadoras para a reindustrialização—PRR, Proposal: C632482988-00467016.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are not available due to privacy concerns.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AIDI	Autonomous Interrogator and Deception Identifier
FER	Facial Expression Recognition
INTU-AI	Intuition Artificial Intelligence
LCS	longest common sequence

PJ	Polícia Judiciária
PJM	Polícia Judiciária Militar
SER	Speech Emotion Recognition
TEA	Text-based Emotion Analysis
WER	Word Error Rate

## References

- Sharma, I.; Kumar, P.; Kaushik, R. Application of AI in Everyday Life. *Ind. Eng. J.* **2022**, *51*, 33–38. [[CrossRef](#)]
- Maqueda, D.C.M. The Data Revolution in Justice. *World Dev.* **2025**, *186*, 106834. [[CrossRef](#)]
- Ejjami, R. AI-Driven Justice: Evaluating the Impact of Artificial Intelligence on Legal Systems. *Int. J. Multidiscip. Res.* **2024**, *6*, 1–29.
- Vagal, M. Understanding The Offender: Behavioral Evidence Analysis In Forensic Interrogation–Methodologies, Emerging Trends And Applications. *Am. J. Psychiatr. Rehabil.* **2025**, *28*, 290–304. [[CrossRef](#)]
- Farber, H.B.; Vyas, A. Truth And Technology: Deepfakes in Law Enforcement Interrogations. *SSRN Electron. J.* **2025**. Available online: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5122595](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5122595) (accessed on 31 July 2025).
- Polícia Judiciária. 2017. Available online: <https://www.policiajudiciaria.pt/> (accessed on 31 July 2025).
- Polícia Judiciária Militar. 2025. Available online: <https://www.defesa.gov.pt/pt/defesa/organizacao/sc/pjm> (accessed on 14 June 2025).
- Bond, C.F., Jr.; DePaulo, B.M. Accuracy of deception judgments. *Personal. Soc. Psychol. Rev.* **2006**, *10*, 214–234. [[CrossRef](#)] [[PubMed](#)]
- Kleinberg, B.; Verschuere, B. How humans impair automated deception detection performance. *Acta Psychol.* **2021**, *213*, 103250. [[CrossRef](#)] [[PubMed](#)]
- Chkroun, M.; Azaria, A. Autonomous Agents for Interrogation. In Proceedings of the 2024 IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI), Herndon, VA, USA, 28–30 October 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 686–693.
- King, S.L.; Neal, T. Applications of AI-Enabled Deception Detection Using Video, Audio, and Physiological Data: A Systematic Review. *IEEE Access* **2024**, *12*, 135207–135240. [[CrossRef](#)]
- Brennen, T.; Magnussen, S. Lie detection: What works? *Curr. Dir. Psychol. Sci.* **2023**, *32*, 395–401. [[CrossRef](#)]
- Dow, D.; Jeu, C.; Watkins, S. How Reliable are Polygraph Examinations in Criminal Investigations? An Empirical Assessment. *SSRN Electron. J.* **2024**. Available online: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5202561](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5202561) (accessed on 30 August 2024).
- Brennen, T.; Magnussen, S. The science of lie detection by verbal cues: What are the prospects for its practical applicability? *Front. Psychol.* **2022**, *13*, 835285. [[CrossRef](#)] [[PubMed](#)]
- Constâncio, A.S.; Tsunoda, D.F.; Silva, H.d.F.N.; Silveira, J.M.d.; Carvalho, D.R. Deception detection with machine learning: A systematic review and statistical analysis. *PLoS ONE* **2023**, *18*, e0281323. [[CrossRef](#)] [[PubMed](#)]
- Rahayu, Y.D.; Faticah, C.; Yuniarti, A.; Rahayu, Y.P. Advancements and Challenges in Video-Based Deception Detection: A Systematic Literature Review of Datasets, Modalities, and Methods. *IEEE Access* **2025**, *13*, 28098–28122. [[CrossRef](#)]
- Patel, K.; Airen, P.; Singh, S. Advance Deception Detection using Multi-Modal Analysis. In Proceedings of the 2025 2nd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 7–8 May 2025; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6.
- Darwin, C. *The Expression of the Emotions in Man and Animals*; John Murray: London, UK, 1872. Available online: <https://www.gutenberg.org/ebooks/1227> (accessed on 14 June 2025).
- Dalgleish, T.; Power, M. *Handbook of Cognition and Emotion*; John Wiley & Sons: Hoboken, NJ, USA, 2000.
- Plutchik, R. Emotion. In *Emotion, A Psychoevolutionary Synthesis*; Longman Higher Education: Harlow, UK, 1980.
- Burgoon, J.K. Microexpressions are not the best way to catch a liar. *Front. Psychol.* **2018**, *9*, 1672. [[CrossRef](#)] [[PubMed](#)]
- Ekman, P. *Telling Lies*; WW Norton: New York, NY, USA, 2002.
- Carson, T.L. *Lying and Deception: Theory and Practice*; Oxford University Press: Oxford, UK, 2010.
- Kulhman, M.S. Nonverbal communications in interrogations. *FBI L. Enforc. Bull.* **1980**, *49*, 6.
- DePaulo, B.M.; Lindsay, J.J.; Malone, B.E.; Muhlenbruck, L.; Charlton, K.; Cooper, H. Cues to deception. *Psychol. Bull.* **2003**, *129*, 74. [[CrossRef](#)] [[PubMed](#)]
- Vrij, A. *Detecting Lies and Deceit: Pitfalls and Opportunities*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
- Cutuli, D. Cognitive reappraisal and expressive suppression strategies role in the emotion regulation: An overview on their modulatory effects and neural correlates. *Front. Syst. Neurosci.* **2014**, *8*, 175. [[CrossRef](#)] [[PubMed](#)]
- Câmara Municipal de Vagos. Gerar PDF Cartão Cidadão. 2025. Available online: [https://www.cm-vagos.pt/cmvgos/uploads/document/file/4995/gerar\\_pdf\\_cartao\\_cidadao.pdf](https://www.cm-vagos.pt/cmvgos/uploads/document/file/4995/gerar_pdf_cartao_cidadao.pdf) (accessed on 14 June 2025).

29. Código de Processo Penal. Diário da República, 1.<sup>a</sup> Série—N.º 40/1987, 1987. Portugal. Available online: <https://diariodarepublica.pt/dr/legislacao-consolidada/decreto-lei/1987-34570075> (accessed on 14 June 2025).
30. OpenAI. Whisper: Open-Source Speech Recognition. 2022. Available online: <https://openai.com/index/whisper/> (accessed on 12 December 2024).
31. Radford, A.; Kim, J.W.; Xu, T.; Brockman, G.; McLeavey, C.; Sutskever, I. Robust speech recognition via large-scale weak supervision. In Proceedings of the International Conference on Machine Learning. PMLR, Honolulu, HI, USA, 23–29 July 2023; pp. 28492–28518.
32. Cristian, R. flan-t5-Portuguese-Small-Summarization. 2025. Available online: <https://huggingface.co/rhaymison/flan-t5-portuguese-small-summarization> (accessed on 14 June 2025).
33. PyMuPDF. Welcome to PyMuPDF. 2025. Available online: <https://pymupdf.readthedocs.io/en/latest/> (accessed on 24 July 2025).
34. Linzaer. Ultra-Light-Fast-Generic-Face-Detector-1MB. 2025. Available online: <https://github.com/Linzaer/Ultra-Light-Fast-Generic-Face-Detector-1MB> (accessed on 2 January 2025).
35. Durai, P. Github—Facial Emotion Recognition. 2023. Available online: <https://github.com/spmallick/learnopencv/tree/master/Facial-Emotion-Recognition> (accessed on 10 November 2024).
36. Calabrés, E.H. wav2vec2-lg-xlsr-en-Speech-Emotion-Recognition. 2025. Available online: <https://huggingface.co/ehcalabres/wav2vec2-lg-xlsr-en-speech-emotion-recognition> (accessed on 14 June 2025).
37. Grosman, J. Hugging Face. 2021. Available online: <https://huggingface.co/jonatasgrosman/wav2vec2-large-xlsr-53-english> (accessed on 10 January 2025).
38. Kapase, A.; Uke, N.; Savant, J.; Desai, M.; Ghatage, S.; Rahangdale, A. “AffectAlchemy”: An Affective Dataset Based on Plutchik’s Psychological Model for Text-Based Emotion Recognition and its Analysis Using ML Techniques. In Proceedings of the 2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA), Pune, India, 23–24 August 2024; pp. 1–6. [CrossRef]
39. DeepL. DeepL Translate. Available online: <https://www.deepl.com/en/translator> (accessed on 13 November 2024).
40. Deep-Translator. 2024. Available online: <https://github.com/nidhaloff/deep-translator> (accessed on 14 November 2024).
41. FastText: Library for Efficient Text Classification and Representation Learning. Facebook Open Source, 2022. Available online: <https://fasttext.cc/> (accessed on 14 June 2025).
42. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
43. Kapase, A. Affect Alchemy: A Dataset for Plutchik-Based Emotion Recognition. 2023. Available online: <https://github.com/ajaykapase/affect-alchemy> (accessed on 4 August 2025).
44. Invideo AI. 2024. Available online: <https://invideo.io/> (accessed on 26 December 2024).
45. Garcia, J.G. Video Corpus for Transcription Testing in the INTU-AI Project. 2025. Available online: <https://drive.google.com/drive/folders/1xgyHinmyZtYfSghvwBGJwI-Jdt-jPaag?hl=pt-br> (accessed on 31 July 2025).
46. kl3z. INTU-IA Multimodal Approach. 2025. Available online: [https://github.com/kl3z/INTU-IA-Multimodal\\_Approach](https://github.com/kl3z/INTU-IA-Multimodal_Approach) (accessed on 31 July 2025).
47. Niesen, J. Jiwer: Speech-Text Evaluation Measures in Python. 2023. Available online: <https://pypi.org/project/jiwer/> (accessed on 31 July 2025).
48. A Sentença. TVI. 2024. [Television Tv Show]. Available online: <https://www.imdb.com/title/tt32119132/> (accessed on 31 July 2025).
49. OpenAI. ChatGPT (GPT-4o). 2025. Available online: <https://openai.com/chatgpt> (accessed on 14 June 2025).
50. DeepSeek. DeepSeek-V2 Language Model. 2025. Available online: <https://deepseek.com/> (accessed on 14 June 2025).
51. Lin, C.Y. Rouge: A package for automatic evaluation of summaries. In Proceedings of the Text Summarization Branches Out, Barcelona, Spain, 25–26 July 2004; pp. 74–81.
52. INTU-AI Project. INTU-AI—Intuition Artificial Intelligence. 2025. Available online: <https://drive.google.com/drive/folders/1U1xONXLtSniIxZP386VrekKvyF0x-FZ?hl=pt-br> (accessed on 31 July 2025).
53. TypeForm. PJM IA Evaluation Questionnaire. Online Questionnaire for Qualitative Evaluation of PJM IA Software. Available online: <https://form.typeform.com/to/YGw8ayvZ> (accessed on 31 July 2025)
54. Likert, R. A Technique for the Measurement of Attitudes. *Arch. Psychol.* **1932**, *22*, 55.
55. Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User Acceptance of Information Technology: Toward a Unified View. *MIS Q.* **2003**, *27*, 425–478. [CrossRef]
56. META. Llama 4. 2025. Available online: <https://www.llama.com/> (accessed on 10 June 2025).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.