Generative AI-based Cognitive Robot for exam candidates' knowledge self-assessment

Intissar Haddiya
Faculty of Medicine and Pharmacy
University Mohamed Premier
Oujda, Morocco
intissarhaddiya@yahoo.fr

Andrea Pitrone
Loop AI Group LLC
New York, United States
andrea@loop.ai

***Abstract* — Medical curricula are based on both theorical knowledge and competence achievement for either postgraduate or undergraduate medical education. It is important that medical students can self-assess their knowledge and competence so that they can take responsibility for their learning quality. Self-assessment is an important parameter in medical education to develop clinical competence.**

In this article we present a Generative AI-based Cognitive Robot developed for enabling the automatic self-assessment of hypertension exam candidates' knowledge.

The candidates are provided with ten direct open-questions automatically generated by the Artificial Intelligence solution (i.e., the Cognitive Robot): the performance of the Cognitive Robot is evaluated by comparing the outcomes calculated by the AI solution and the results achieved by the students while answering to the same list of questions, asked by a human Investigator.

The Cognitive Robot has proven to present a very high level of accuracy. Moreover, the candidates involved in this study have confirmed the usefulness and trust of the Cognitive Robot. The Artificial Intelligence solution proposed for self-assessing the exam candidates' knowledge is effective, innovative, accurate and can be extended to other field of study in the medicine realm.

Keywords* — *Medical students, Self-assessment, Micro-facial expressions, Artificial Intelligence, Generative AI, Cognitive Robot, Unsupervised Learning, Large Language Model (LLM), Convolutional Neural Network (CNN).

I. Introduction

Medical curricula are based on both theorical knowledge and competence achievement for either postgraduate or undergraduate medical education [1] [2]. It is important that medical students can self-assess their knowledge and competence so that they can take responsibility for their learning quality [3]. Self-assessment is an important parameter in medical education to develop clinical competence [4] [5].

The ability to accurately assess one's own competence – and particularly its limits – in combination with feedback from faculty, provides a valuable learning tool encouraging reflection [6] [7] [8] [9]. Moreover, self-assessed knowledge is important for investigating medical students' preparedness for exams [10].

However, medical students' self-assessment is not always accurate, as both under- and overestimation of competence or knowledge has been found in previous studies [11].

Patients' medical diagnoses and treatments are based on per-specific-morbility protocol, composed by a standard and fixed set of steps, recognized, and adopted at a global level.

Therefore, the assessment of medical students' knowledge concerning any specific diagnoses and treatments requires the verification of a repeatable and standard list of items, expressed by means of "technical" and medical topic-related terms and concepts.

Artificial intelligence (AI) has emerged as a promising set of tools and methodologies for assessing a variety of applications, including medical diagnoses and treatments, e.g., for automating and objectifying the detection of human body areas affected by cancer [12].

This article describes the Generative AI Cognitive Robot that we have designed and developed for helping medical students assess their own readiness for exams through a simulation of the exam itself.

II. Cognitive Robot Solution Working Principles

The Cognitive Robot can be used as a mobile application or on desktop computers in order to record a video of every single exam simulation: the video is elaborated by the Cognitive Robot itself and the detailed results with overall outcomes ("Ready" vs. "non-Ready" as well as correct vs. incorrect answers) are then provided to the candidates.

The Cognitive Robot is initially catered with the documentation including the knowledge that the candidates have to study and are expected to acquire: this is called the "training phase" of the Cognitive Robot.

During the exam simulation session, the Cognitive Robot automatically generates ten direct open-questions to be asked to the candidate and records a video of the candidate while answering to the questions during a session lasting about 15-minutes.

Once the recording session is completed, the Cognitive Robot evaluates the correctness and completeness of the answers by referencing the documentation exploited by the student, i.e., the same documentation leveraged initially for the training phase. For each question/answer set, a number of specific topic-related concepts is expected to be mentioned by the candidate, as included in the reference documentation, this one being the criteria for evaluating the completeness of the answer. The student is expected to provide an answer ideally bearing the whole list of concepts or synonyms of the same concepts.

The Cognitive Robot will consider any evaluated answer correct or partially correct if:

- the wording of the answer is semantically similar to the related content in the reference documentation;
- the answer includes the full list of concepts/keywords (or synonyms) – in case a partial list of concepts is detected in the answer, then a percentage of completeness is calculated by the Cognitive Robot.

Eventually, the Cognitive Robot will calculate the final mark of the exam simulation session by considering the evaluation of every single answer provided by the candidate.

For instance, a question generated by the Cognitive Robot might be: "How do you confirm the diagnosis of hypertension?" The correct and complete answer should include the following key concepts: "Because of the variability of blood pressure, an elevation of office BP (SBP $\geqslant$140 mmHg or DBP $\geqslant$90 mmHg) should be <u>confirmed by at least two to three visits, unless the BP values recorded during the first visit are markedly elevated (grade 3 hypertension) or CV risk is high, including the presence of HMOD</u>" (the relevant concepts to be mentioned by the candidates in order to consider the answer correct and complete are underlined).
The following response is considered correct and complete as it includes all main concepts: "Due to the phenomenon of blood pressure variability, an elevation of office BP (SBP $\geqslant$ 140 mmHg or DBP $\geqslant$90 mmHg) need to be verified and ascertained by means of a minimum number of visits equal to two or three, unless the values of blood pressure calculated while the patient was having the first visit are clearly high likewise in the grade 3 of hypertension or else the cardiovascular risk is relevant, including the presence of mediated organ damage".
However, the following answer is considered correct but incomplete, as some relevant concepts are missing: "Due to the phenomenon of blood pressure variability, an elevation of office BP (SBP $\geqslant$140 mmHg or DBP $\geqslant$90 mmHg) need to be verified and ascertained by means of a minimum number of visits equal to two or three".

Furthermore, the Cognitive Robot detects and evaluates the micro-facial expressions of the candidate while answering every single question, with the objective of taking into consideration the reliability and self-confidence of the student with every topic.

The reference methodology adopted for evaluating the micro-facial expressions relies on Paul Ekman's studies [13] [14]. Paul Ekman discovered that some facial expressions of emotion are universal while many of the apparent differences in facial expressions across cultures were due to context. He also co-discovered micro facial expressions. The seven universal facial expressions are: Happiness, Sadness, Fear, Disgust, Anger, Contempt and Surprise.

For this study, the Cognitive Robot has been developed for dealing with hypertension-related knowledge; nevertheless, its adoption in the Medicine Universities can be easily generalized for covering any other medical field of application.

III. Methodology Applied For Our Study

A. Main Methodological Criteria

Our Study has been executed by applying the criteria represented as follows:

- <u>Participants</u>: Fifty fifth-year medical Students who were highly motivated to take part to the study have been involved, providing their own consent to participate.
- <u>Type of study</u>: A prospective, descriptive, and analytical study, conducted in the department of Nephrology in the Mohammed VI University Hospital of Oujda-Morocco.
- <u>Field of application</u>: Hypertension comprehensive data.

The students have been provided with an about 500-pages document covering information in the diagnosis and management of hypertension, as well as principles of hypertension pathophysiology based on recent hypertensive guidelines and trials. This activity is based on the European guidelines of hypertension in 2023.

The students have been given a time period of ten days to study the document. At the conclusion of this initial activity, participants were able to:

- Perform appropriate diagnostic assessments;
- Identify pathophysiologic factors of high-blood pressure and their treatment implications;
- Provide appropriate prescription regimens in resistant hypertension-affected patients;
- Apply recent guidelines and studies to hypertensive care.

B. Assessment and Results Evaluation

The Cognitive Robot solution effectiveness has been measured by comparing the outcomes of the candidates' knowledge evaluation executed by exploiting two different methods:

- Cognitive Robot-based exam simulation, which involves students utilizing the functionalities of the Cognitive Robot and responding to ten automatically generated open-ended questions provided by the Artificial Intelligence solution;
- Human Investigator-based assessment, which involves evaluating candidates' readiness through a 15-minute individual oral questionnaire session, wherein the same questions generated by the Cognitive Robot are asked.

As a final step of the study, the outcomes of the Cognitive Robot have been compared with the ones evaluated by the Human Operator, in order to validate the AI technical solution in terms of candidates' readiness assessment accuracy, while the most trusted evaluation results has been considered the ones made by the Human Investigator, taken as a reference. Therefore, in order to evaluate the Cognitive Robot solution accuracy, we have calculated the Mean Squared Error of the Exam Readiness percentage calculated by means of the Cognitive Robot and Human Investigator methods.

IV. Cognitive Robot Solution Components

The Cognitive Robot is composed by two main AI components:

- Natural Language Engine ("NLE"), for natural language understanding, processing, and generation;
- Multimodal Sentiment Analyzer ("MSA"), for detecting emotions from the exam candidates' video by exploiting:
  - microfacial expressions from images/frames and whole video;
  - audio nuance;
  - text (i.e., from the speech-to-text conversion).

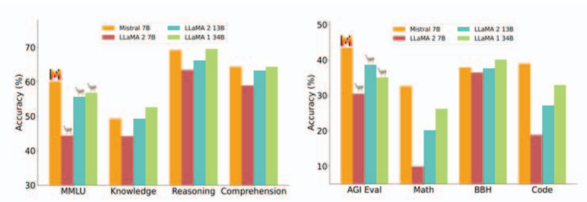
A. *Natural Language Engine*

The Natural Language Engine ("NLE") consists of an open-source LLM (large language model) fine-tuned with specific hypertension-related content and specifically by leveraging the document provided to the students, from which the exam simulation questions are automatically generated by the Cognitive Robot. The total training duration for fine-tuning this model amounted to 10 hours.

We adopted Mistral 7B, a 7.3B parameters Large Language Model (LLM) based on a Transformer architecture and engineered for superior performance and efficiency. It leverages grouped-query attention (GQA) for faster inference, coupled with sliding window attention (SWA) to effectively handle sequences of arbitrary length.

Mistral 7B outperforms the best open 13B model (Llama 2) across all evaluated benchmarks, and the best released 34B model (Llama 1) in reasoning, mathematics, and code generation, as represented on Fig. 1, comparing the performance of several Large Language Models (LLM).

Fig. 1. Comparison of main LLMs performance metrics values



During the exam simulation session, the NLE randomly generates ten questions to ask to the candidate, by extracting content from the document studied by the candidate. The student answers every single question, while the tool records the video, in order to subsequently generate basic components for text, audio, video and images/frames.

The Cognitive Robot is able to evaluate the correctness of every single answer, by comparing the one provided by the candidate with the content of the aforementioned document. We have leveraged the Loop AI proprietary unsupervised learning platform to create the knowledge model in the Tool, so that the Cognitive Robot can "understand" if the answer is correct despite the usage of synonyms and/or different wording in the candidate's answer. The Unsupervised Learning approach also allows to properly handle any potential misspellings generated by the Speech-to-Text module, adopted to convert the audio file to text [15].

In particular, during the "Learning phase", the Unsupervised Platform is catered with a set of context-related documentation / data (i.e., hypertension, for our study): the Platform is able to automatically create a contextual representation of the main concepts included within the training dataset and identify the semantic correlation of main concepts and sub-concepts, including the "misspelt versions" of main relevant terms.

The correctness percentage of every single answers provided by the candidates to the ten questions in the exam simulation session is calculated by detecting the number of keywords (or related synonyms) included in the answer itself as compared to the reference document [19], as a measure of the completeness of the answer. The human investigator employs identical criteria when assessing the candidates' responses to the oral questionnaire, which consists of the same questions generated by the AI solution.

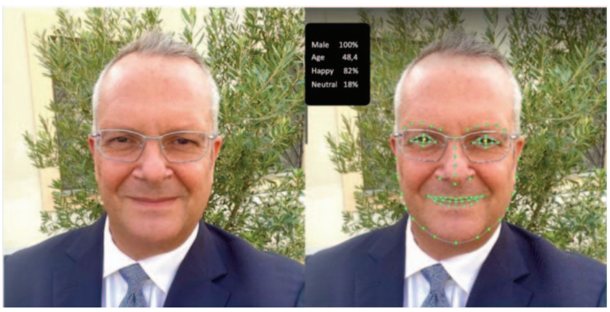
Furthermore, the Cognitive Robot evaluates the semantic similarity of the candidates' answers to the related content for the specific question as indicated in the reference document [19]. Eventually, the Robot indicates the expected correct answer for every single question, by extracting the related content (text and list of keywords/concepts) from the document studied by the candidates.

Therefore, the AI solution sets forth an "Explicable AI" approach, thus providing a justification for the evaluation of the answer's correctness and completeness.

B. *Multimodal Sentiment Analyzer*

The Multimodal Sentiment Analyzer "MSA" module is a CNN-based component, capable to detect the seven types of emotion identified by Paul Ekman [13] [14] [16] [17] [18]: Happiness, Sadness, Fear, Disgust, Anger, Contempt and Surprise, as shown on Fig. 2, providing an example of facial expressions' analysis.

Fig. 2. Example of facial expressions' analysis



The model has been trained by exploiting data from the Extended Cohn-Kanade dataset as well as own custom images and pics. Also, some pre-processing tasks have been leveraged for improving the overall accuracy of the model. The total training duration for this model amounted to approximately 22 hours.

The development of the MSA component has required the execution of the following subtasks:

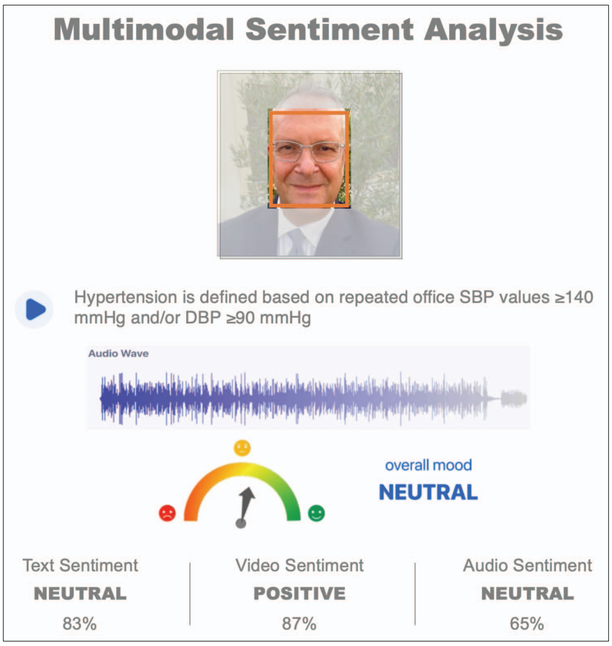
- Text Sentiment Analyzer:
    - Speech-to-Text Converter: we exploited a Loop AI proprietary algorithm which receives an .wav audio file as input and extracts the text as output;
    - Sentiment Analyzer: we exploited a Loop AI proprietary algorithm which classifies utterances as “positive”, “negative” or “neutral” and provides a confidence level percentage for every single classification task executed.
- Audio Sentiment Analyzer: we exploited a Loop AI proprietary algorithm which receives an .wav audio file and executes the related classification as “positive”, “negative” or “neutral” and provides a confidence level percentage for every single classification task executed.
- Video Sentiment Analyzer
    - Dataset creation: we selected a number of labeled facial expressions from different sources and converted the original labels to one of the Ekman’s expression types.
    - Images pre-processing: we chose an open-source software capable to identify the faces section within an image, despite the orientation and/or rotation of a person within a recorded video;
    - CNN development:
        - we trained a multilayer CNN for classifying the facial expression starting from a static cropped image;
        - we adopted VGG16 as CNN;
        - starting from the available dataset, we applied a 70% / 20% / 10% pattern respectively for the training / test / validation activities of the CNN;
        - during the validation phase we have detected an accuracy of 91.3% in the facial expressions’ classification;
    - Video Analyzer development: we chose a software tool for elaborating a video stream and extracting cropped static images from the video itself, for catering the submodules aforementioned.

A User Interface enables the visualization of the sentiment analysis outcome, represented as:

- Overall sentiment;
- Text Sentiment;
- Audio Sentiment;
- Video Sentiment,

as indicated in Fig. 3, showing an example of Multimodal Sentiment analysis.

Fig. 3. Example Multimodal Sentiment analysis



V. Cognitive Robot Technical Infrastructure

The Cognitive application is hosted on redundant on-premises infrastructure, consisting of two Servers HPE DL380 Gen10 configured in an active-standby setup, each equipped as follows:

- CPU: 2x Intel Xeon Gold 6130 2.1 GHz, 3,7 GHz Turbo, 16C, 22,75 MB di cache L3 (125 W) DDR4-2666
- MEM: 64 GB (2 x 32 GB) Rx4 PC4-2666V-R
- Operating System: Linux Centos 7.x
- GPU: 2 x Tesla V100 32GB

The Cognitive application also supports deployment on either a private or public cloud infrastructure.

Each video, comprising the entire set of 10 open-ended questions, lasted approximately 5-8 minutes. The Cognitive Robot required about 10-15 minutes to process the video and provide the results as mentioned earlier.

With this infrastructure configuration, it's feasible to simultaneously process a maximum of two videos or exam simulation sessions (in other words, one video occupies one GPU for the entire processing duration).

VI. Results Achieved

In this section, we analyze the outcomes from two separate sessions conducted with each individual student. The first session utilizes the capabilities of the Cognitive Robot, while the second one is led by a Human Investigator.

The "Exam Readiness threshold" is configurable in the Cognitive Robot software application and it has been initially set to 80% as minimum value. Therefore, in case the Exam Readiness% calculated value is lower than 80%, then the Student should strongly consider postponing the exam session, whereas a value higher than 80% means that the Student is ready for the exam.

The "Exam Readiness %" value is calculated either in case of session executed with the Cognitive Robot or oral questionnaire with the human Investigator as the average value of the sum of the percentages achieved in each one of the ten open-questions.

The comparison of the results achieved by means of the two methods for simulating the exam session (i.e., session executed with the Cognitive Robot or oral questionnaire with the human Investigator) is represented in the Table I below.

TABLE I. COMPARISON OF ACHIEVED RESULTS

ID	*Student*	Exam Readiness %		*Mean Squared Error (MSE) calculation*
		Human Investigator (y)	*Cognitive Robot (p)*	
1	***Student 1***	***80%***	***75% (*)***	***$(y\text{-}p)^2 = 0{,}0025$***
2	Student 2	80%	82%	***$(y\text{-}p)^2 = 0{,}0004$***
3	Student 3	100%	98%	***$(y\text{-}p)^2 = 0{,}0004$***
4	Student 4	60%	63%	***$(y\text{-}p)^2 = 0{,}0009$***
5	Student 5	50%	51%	***$(y\text{-}p)^2 = 0{,}0001$***
6	Student 6	80%	84%	***$(y\text{-}p)^2 = 0{,}0016$***
7	Student 7	100%	100%	***$(y\text{-}p)^2 = 0{,}0000$***
8	Student 8	50%	48%	***$(y\text{-}p)^2 = 0{,}0004$***
9	Student 9	60%	62%	***$(y\text{-}p)^2 = 0{,}0004$***
10	Student 10	80%	83%	***$(y\text{-}p)^2 = 0{,}0009$***
11	Student 11	70%	68%	***$(y\text{-}p)^2 = 0{,}0004$***
12	Student 12	90%	92%	***$(y\text{-}p)^2 = 0{,}0004$***
13	***Student 13***	***80%***	***77% (*)***	***$(y\text{-}p)^2 = 0{,}0009$***
14	Student 14	60%	62%	***$(y\text{-}p)^2 = 0{,}0004$***
15	Student 15	100%	98%	***$(y\text{-}p)^2 = 0{,}0004$***
16	Student 16	100%	100%	***$(y\text{-}p)^2 = 0{,}0000$***
17	Student 17	40%	37%	***$(y\text{-}p)^2 = 0{,}0009$***
18	Student 18	100%	98%	***$(y\text{-}p)^2 = 0{,}0004$***
19	Student 19	90%	93%	***$(y\text{-}p)^2 = 0{,}0009$***
20	Student 20	80%	84%	***$(y\text{-}p)^2 = 0{,}0016$***
21	Student 21	100%	100%	***$(y\text{-}p)^2 = 0{,}0000$***
22	Student 22	60%	57%	***$(y\text{-}p)^2 = 0{,}0009$***
23	***Student 23***	***80%***	***78% (*)***	***$(y\text{-}p)^2 = 0{,}0004$***
24	Student 24	60%	63%	***$(y\text{-}p)^2 = 0{,}0009$***
25	Student 25	100%	97%	***$(y\text{-}p)^2 = 0{,}0009$***
26	Student 26	90%	93%	***$(y\text{-}p)^2 = 0{,}0009$***
27	Student 27	70%	73%	***$(y\text{-}p)^2 = 0{,}0009$***
28	Student 28	60%	58%	***$(y\text{-}p)^2 = 0{,}0004$***
29	Student 29	100%	100%	***$(y\text{-}p)^2 = 0{,}0000$***
30	Student 30	60%	62%	***$(y\text{-}p)^2 = 0{,}0004$***
31	Student 31	70%	67%	***$(y\text{-}p)^2 = 0{,}0009$***
32	Student 32	100%	100%	***$(y\text{-}p)^2 = 0{,}0000$***
33	Student 33	40%	37%	***$(y\text{-}p)^2 = 0{,}0009$***
34	Student 34	80%	84%	***$(y\text{-}p)^2 = 0{,}0016$***
35	Student 35	100%	100%	***$(y\text{-}p)^2 = 0{,}0000$***
36	Student 36	100%	98%	***$(y\text{-}p)^2 = 0{,}0004$***
37	Student 37	40%	37%	***$(y\text{-}p)^2 = 0{,}0009$***
38	Student 38	100%	97%	***$(y\text{-}p)^2 = 0{,}0009$***
39	Student 39	50%	52%	***$(y\text{-}p)^2 = 0{,}0004$***
40	***Student 40***	***80%***	***76% (*)***	***$(y\text{-}p)^2 = 0{,}0016$***
41	Student 41	100%	100%	***$(y\text{-}p)^2 = 0{,}0000$***
42	Student 42	90%	91%	***$(y\text{-}p)^2 = 0{,}0001$***
43	Student 43	60%	58%	***$(y\text{-}p)^2 = 0{,}0004$***
44	Student 44	50%	46%	***$(y\text{-}p)^2 = 0{,}0016$***
45	***Student 45***	***80%***	***75% (*)***	***$(y\text{-}p)^2 = 0{,}0025$***
46	Student 46	100%	97%	***$(y\text{-}p)^2 = 0{,}0009$***
47	Student 47	60%	63%	***$(y\text{-}p)^2 = 0{,}0009$***
48	Student 48	80%	83%	***$(y\text{-}p)^2 = 0{,}0009$***
49	Student 49	40%	43%	***$(y\text{-}p)^2 = 0{,}0009$***
50	Student 50	100%	100%	***$(y\text{-}p)^2 = 0{,}0000$***
			MSE	**0,07%**

(*) "False negative" predictions out of 50, since the "Readiness threshold" was 80%

The two methods have provided the same results in terms of exam readiness evaluation, apart from Students 1, 13, 23, 40 and 45 cases, for which the two methods provided different outcomes (i.e., "ready" for the Human Investigator approach as opposed to "not ready" for the Cognitive Robot). Nevertheless, in these specific cases, a proper fine-tuning of the threshold value (currently set to 80%) could also remove this "false negative" result and raise-up the accuracy level of the Cognitive Robot to the value of 100%.

VII. CONCLUSIONS

The Cognitive Robot has proven to present a very high level of accuracy (90% - only five evaluation errors out of 50 cases), by comparing the calculated results and the ones evaluated by the Human Investigator.

Also, a Customer Satisfaction survey has been filled in by the involved students, confirming the usefulness and trust of the Cognitive Robot. This means that the students would be willing to adopt the automated Cognitive Robot-based solution also for other topics/fields of study within the medicine realm, which turns out to be a successful outcome for our study, given that the will to adopt and exploit an AI-based automated solution replacing the human-to-human interaction experience is the most relevant challenge to cope with and address right from the start of the Cognitive Robots solution design.

As a prospective expansion of this study, we intend to increase the number of candidates under analysis. Additionally, we plan to employ the same Cognitive Robot across various medical domains, such as kidney disease treatments, to reaffirm the broader applicability and precision of the AI solution.

This study was carried out in compliance with the Helsinki declaration on the protection of individuals and was approved by the Ethics Committee of the Faculty of Medicine and Pharmacy- University Mohammed Premier, Oujda (Morocco). Students' participation was voluntary. All participants were informed of the objectives and procedures of the study. There was no request for personal information and the confidentiality of the data was guaranteed.

REFERENCES

[1] P. Harris, L. Snell, M. Talbot, RM. Harden, "Competency-based medical education: implications for undergraduate programs", Med. Teach. 2010, 32(8):646–650.

[2] J. Pimentel, J. Garcia, A. E. Romero Tapia, G. Zuluaga, C. A. Correal-Muñoz, A. Cockcroft, N. Andersson, "Competency-Based Cultural Safety Training in Medical Education at La Sabana University, Colombia: A Roadmap of Curricular Modernization", Teaching and Learning in Medicine 2023, 1–10.

[3] S. Prediger, K. Schick, F. Fincke, S. Fürstenberg, V. Oubaid, M. Kadmon, PO. Berberat, S. Harendza, "Validation of a competence-based assessment of medical students' performance in the physician's role", BMC Med. Educ. 2020, 20:6.

[4] M. Wijnen-Meijer, M. van der Schaaf, K. Nillesen, S. Harendza, O. Ten Cate, "Essential facets of competence that enable trust in graduates: a Delphi study among physician educators in the Netherlands", J Grad. Med. Educ. 2013, 5(1):46–53.

[5] K. Levandovska, "Formation of professional competence in medical students in distance education conditions", Bukovinian Medical Herald 2022, 26(4 (104)):82-88.

[6] D. Blanch-Hartigan, "Medical students' self-assessment of performance: results from three meta-analyses", Patient Educ. Couns. 2011, 84(1):3–9.

[7] S. Shrestha, S. Gurung, B. Bhandari, "Self-Assessment of Undergraduate Medical Students Regarding the Clinical Knowledge in Forensic Medicine", Journal of Nepalgunj Medical College 2022, 20(2):10-13.

[8] RM. Epstein, "Assessment in medical education", N. Engl. J Med. 2007, 356:387–396.

[9] M. N. N. Htay, G. Kamath, A. KM, S. Sugathan, E. O. Luyee, S. Sahoo, "Overview of Reliability and Validity of Assessments in Medical Education", International Journal of Transformative Health Professions Education 2023, Vol. 1 19-26.

[10] M. Mette, C. Baur, J. Hinrichs, E. Narciß, "Gaining interprofessional knowledge and interprofessional competence on a training ward", Med. Teach. 2021, 43(5):583–589.

[11] N. Oosterom, L. Floren, O. ten Cate, H. Westerveld, "A review of interprofessional training wards: enhancing student learning and patient outcomes", Med. Teach. 2019, 41(5):547–554.

[12] M. Cesarelli, F. Mercaldo, A. Santone, "Explainable Deep Learning for Brain Cancer Detection and Localisation", IEEE MEtroXRaine 2023, Proceedings: 496-501.

[13] K. Ramsland, "The man of 1,000 faces: Paul Ekman and the science of facial analysis", Springfield Vol. 22, Iss. 1, (Winter 2012): 65-70.

[14] D. Schiller, "The face and the faceness: Iconicity in the early faciasemiotics of Paul Ekman, 1957–1978", Sign Systems Studies 2021, 361–382.

[15] E. Schumacher, M. Dredze, "Learning unsupervised contextual representations formedical synonym discovery", AMIA Open, 2(4), 2019, 538–546.

[16] T. Gilligan, B. Akis, "Emotion AI, Real-Time Emotion Detection using CNN", https://web.stanford.edu/class/cs231a/prev_projects_2016/emotion-ai-real.pdf.

[17] S. Bekenova, A. Bekenova, "Emotion recognition and classification based on audio data using AI", E3S Web of Conferences 2023, 420. 10.1051/e3sconf/202342010040.

[18] J. A. Ballesteros, G. M. Ramirez, F. Moreira, A. Solano, C. A. Pelaez, "Facial emotion recognition through artificial intelligence", Frontiers in Computer Science 2024, 6. 10.3389/fcomp.2024.1359471.

[19] G. Mancia et alii, "2023 ESH Guidelines for the management of arterial hypertension. The Task Force for the management of arterial hypertension of the European Society of Hypertension Endorsed by the European Renal Association (ERA) and the International Society of Hypertension (ISH): Journal of Hypertension".