I. IDENTIFICATION DATA

Thesis name:	**Sophia the Robot - Development of a Software Extension - Detection and Reaction to Being Photographed by a Phone Camera**
Author's name:	**Jakub Šura**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	Doc. Mgr. Matěj Hoffmann, PhD
Reviewer's department:	Department of Cybernetics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment	**challenging**

Evaluation of thesis difficulty of assignment.

The assignment involved adding a new functionality to a humanoid robot, extending and integrating into existing perceptual and behavioral modules.

Satisfaction of assignment	**fulfilled with minor objections**

Assess that handed thesis meets assignment. Present points of assignment that fell short or were extended. Try to assess importance, impact or cause of each shortcoming.

The outcome of the thesis is working software implementing the desired functionality. However, the evaluation of the new addition seems preliminary and I could not find "The success will be rated by objective metrics (e.g., response time, gaze accuracy) and subjective metrics (e.g., how good does the picture look)."

Method of conception	**partially applicable**

Assess that student has chosen correct approach or solution methods.

The project required the design and integration of additions to different components or layers of the overall robot architecture – perception, attention, reasoning, and behavior. Appropriate addition of the "someone is taking a picture of me" module to each of them is not easy. The student has partially taken a pragmatic "ad hoc" approach to "get the things done" and selected solutions that were easily available. For example, the YOLOv8-Pose method for human pose detection does not have state-of-the-art performance.

Technical level	**B - very good.**

Assess level of thesis specialty, use of knowledge gained by study and by expert literature, use of sources and data gained by experience.

The student has delivered the "someone is taking a picture of me" functionality into the robot architecture, which is a remarkable accomplishment. However, as mentioned under "Method of conception" above, the choice of the solutions for the individual functionalities seems partially arbitrary and based on convenience and "internet discussions". I can illustrate it on the human pose estimation model (pg. 18-20). The reference used about existing solutions is a "random blog" [31] and the method chosen (YOLOv8-Pose) is not state of the art.

"…pose estimation model Pose-ResNet18-Body capable of detecting up to 18 keypoints of human bodies (as described by Lin,Maire, et al. in [4]) in real-time" [4] seems to be a reference to an article about the COCO dataset.

There are much better performing methods (ViTPose, AlphaPose, HRNet TD, … according to the standard metrics – OKS, AP, AR), some of which can also satisfy the requirements to run online.

The evaluation of the developed solution is preliminary.

Formal and language level, scope of thesis	**C - good.**

Assess correctness of usage of formal notation. Assess typographical and language arrangement of thesis.

The scope of the thesis is appropriate and the thesis reads quite well.

The abstract is too high-level / general. It could better summarize what has been specifically done. The Czech abstract is not perfect. There are stylistic and grammatical errors (missing commas) in the first paragraph.

The student is relying on texts from Hanson robotics that have a product advertisement character. E.g.,:

"Sophia the Robot, developed by Hanson Robotics Limited, is recognized for her state-of-the-art human-like expressive face 1.1 and for being one of the first robots of her kind to achieve celebrity status. Unveiled in 2016 [5], Sophia has become an icon in the realm of social robotics, distinguished by her ability to mimic human emotions and engage in meaningful dialogues. Her sophisticated appearance and advanced AI integration make her a key figure in exploring the intersection of technology and social interaction [6]." (should also say Fig. 1.1)

There are more occasions where the student's description is not objective and uses for example "sophisticated" to characterize some features. This is not appropriate here.

A schematics of the whole software architecture would be useful.
In Fig. 2.1, some important features like the neck DoFs and the cameras in the eyes are not marked.

The formatting of "quotes" and dashes ("-" used where it should be n-dash "–" or m-dash "—").

Selection of sources, citation correctness   **D - satisfactory.**

Present your opinion to student's activity when obtaining and using study materials for thesis creation. Characterize selection of sources. Assess that student used all relevant sources. Verify that all used elements are correctly distinguished from own results and thoughts. Assess that citation ethics has not been breached and that all bibliographic citations are complete and in accordance with citation convention and standards.

The list of references contains 68 entries which is a large number. However, most of these sources are links to online blogs, popular science articles, product sites or software repositories.
Only few articles are from reputable venues.
Bibliography should not be an Appendix. Some references are incomplete – for example: [18]. Within the text, it is customary to use only FirstAuthor et al. [REF].
For example, here are some references that I would recommend regarding gaze control for humanoid robots and human-robot interaction:

- Roncone, A., Pattacini, U., Metta, G., & Natale, L. (2016). A Cartesian 6-DoF Gaze Controller for Humanoid Robots. Robotics: Science and Systems, 2016.
- Lehmann, H., Keller, I., Ahmadzadeh, R., & Broz, F. (2017). Naturalistic Conversational Gaze Control for Humanoid Robots—A First Step. In A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyssel, & H. He (Eds.), Social Robotics (pp. 526–535). Springer International Publishing.
- Stanton, C. J., & Stevens, C. J. (2017). Don't stare at me: The impact of a humanoid robot's gaze upon trust during a cooperative human–robot visual task. International Journal of Social Robotics, 9, 745–753.

Additional commentary and evaluation
Present your opinion to achieved primary goals of thesis, e.g. level of theoretical results, level and functionality of technical or software conception, publication performance, experimental dexterity etc.

III. OVERALL EVALUATION, QUESTIONS FOR DEFENSE, CLASSIFICATION SUGGESTION

Summarize thesis aspects that swayed your final evaluation. Please present apt questions which student should answer during defense.

The solution developed by the student is not properly evaluated. There are no objective metrics (the assignment (e.g., response time, gaze accuracy) and the evaluation by using the questionnaire is preliminary. There were 15 employees of the company asked to evaluated the behavior of the robot. However, there is no baseline

(comparison to the state before) and it is also a highly biased group of participants. Naïve subjects should be recruited instead.

Overall, the student accomplished a lot in this thesis – delivered a complex piece of software integrated into a humanoid robot. However, sometimes "less is more". Some of the choices made seem not well grounded in the literature and the formal quality of the final document is not perfect.

I evaluate handed thesis with classification grade **C - good.**

Questions for the defense:
1. The directional microphone array (Fig. 2.5) you propose for sound source localization – what are its dimensions and how would you integrate it into the robot? Human-like appearance of the robot is a critical feature….
2. Torso yaw (Section 2.1.3, pg. 13-14, Fig. 2.6). You need the torso yaw DoF. However, there are 6 DoF in the picture. Why?

Date: **7.6.2024**   Signature: