International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2386-2392
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0320

Multilingual Text-Based Augmentative Communication Using Eye Tracking and Predictive Language Models

Naveen G G[1], Priyadharshan B[2], MS K. Dhanabhavithra[3]

[1,2]UG - Artificial Intelligence and data science, St. Joseph's college of engineering, OMR, Chennai, 600119, India.

[3]Assistant Professor, Artificial Intelligence and data science, St. Joseph's college of engineering, OMR, Chennai, 600119, India.

Emails: *naveengg743@gmail.com[1], dharshanofflac24@gmail.com[2], dhanabhavithra@gmail.com[3]*

Abstract

Effective human-computer interaction remains a major challenge for people with severe motor impairments since most of them have very limited access to conventional input devices like keyboards and mice. Augmentative and Alternative Communication systems address the challenge; however, many of the existing solutions are expensive, complicated, and linguistically restrictive. The present work proposes a low-cost, text-based AAC platform that effectively integrates webcam-based eye and head movement tracking with intelligent language prediction to enhance communicational efficiency. The system utilizes techniques from computer vision for estimating gaze direction and head pose and allows for virtual keyboard-based text entry without using any special hardware. In this work, to further improve the typing speed and reduce input errors, a neural language model generates context-aware word suggestions during the text composition process. Unlike most existing AAC solutions that are bound to a single language, the proposed framework supports multilingual communication by translating the generated text into the user's preferred language, hence extending the accessibility for diverse linguistic populations. The system's performance is evaluated in terms of typing performance, prediction accuracy, and adaptability; it stands out as an effective, low-resource, and inclusive communication solution for users with significant motor disabilities.

Keywords: *Index Terms—AAC, Eye Tracking, Multilingual Language Models, Predictive Text, CNN, Assistive Technology, Multilingual Tokenization, Context-Aware Recommendation.*

1. Introduction

Traditionally, human-computer interaction occurs with the help of input devices like keyboards and mouse. These interactions require a high degree of motor coordination and physical movement. Although these interfaces are successful for many people, they pose significant difficulties for people with serious motor-related disabilities. Therefore, to improve human-computer interactions and address these difficulties, human-computer interactions have investigated gesture-based interactions and speaking interfaces and have developed assistive techniques with the help of sensors. These techniques play an important role in improving mobility and supporting social interactions for people with disabilities. Augmentative and Alternative Communication, abbreviated as AAC, is an important aspect of technology in many ways. People who cannot talk or produce words use AAC to express themselves without making any sounds. The first technology in AAC was primitive, with all devices in this class having a board with a single symbol. The user then points to a picture to convey a message. Computers with audible speech came later with digital electronics in this technology class. The computers produce audible speech when pressed by a finger on a screen. The display flips a new page when pressed by a user who needs a new page. AAC is a branch of assistive technology. All current systems in this

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2386-2392
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0320

technology class have four units. The first is a group of symbols, with these symbols being either alphabetical or having a picture. The second is hardware, which is where the software storing or creating a symbol resides. The third unit is software, which involves the storage and creation of symbols. The final component is the set of strategies that enhance fast communication with minimal chances of error. Cheap webcams and phones have let more people use digital AAC systems. For those who can move only their eyes, eye tracking gear has become the main way to steer the software. A camera notes where the pupil points plus that spot picks a letter on the screen keyboard. Special cameras and infrared hardware keep the price of commercial eye tracking systems high. The equipment is bulky, reacts to changes in light plus loses accuracy when the user moves his head. Operators often must recalibrate the device - routine use becomes awkward. Computer vision research and new AI models now let a standard webcam follow gaze. Convolutional neural networks locate facial landmarks but also calculate where the eyes look without extra sensors. However, at the same time, it was possible to improve the AAC systems regarding text predictions by utilizing natural language processing and larger language models, which facilitated fast and effective communication. In a nut- shell, the use of language predictions in AAC systems aimed at reducing the number of selections needed to construct a proper sentence. Additionally, value was added to the AAC systems through keyboards, feedback, and correction. However, the affordability of these AAC devices seems unattainable for people with severe motor disabilities due to cost issues, as well as due to the unavailability of inexpensive interfaces. To overcome the drawbacks that have been discussed in this paper, an inexpensive AI-based multilingual AAC device has been designed, which will also enable the usage of webcams to track eye/head movements. This will enable the people with severe motor disabilities, who use AAC devices, to use a virtual keyboard with the help of eye/head movements. A predicting model has also been incorporated, which will predict the text to be typed by certain contexts. A significant novelty in the proposed framework is the multilingual support thereof. While initially the text is produced in English, it is easily machine-translatable to the user's language of choice. The support of such multilingualism will also bear the general spread of usability across linguistic diversity. Adaptive interfaces, Unicode rendering, and text-to-speech output are all fully integrated into the system for smooth communication in multiple languages. The remainder of this paper is organized as follows: Section II presents a discussion on relevant literature and existing AAC systems with their characteristics and drawbacks. Section III is devoted to a discussion on the methodology and architectural design of proposed systems. Section IV presents experimental results with performance analysis. Finally, in Section V, this paper is concluded with a glance at important findings and future research in this area [1-10].

2. Literature Review

The researcher, in the literature reviews section, will look for different studies on eye-controlled interaction devices, especially focusing on using virtual keyboards for people suffering from motor impairments. There have been significant advances in Augmentative and Alternative Communication devices, helping people communicate effectively through computational vision and deep learning technologies. Convergence of eye-tracking, AI, and natural language processing has been observed. In the research work presented by authors Waideman and Aquino Junior (2025), titled" AAC System Based on Real- Time Eye Movement Detection with Virtual Keyboard," the possibility of designing a system of AAC with a detection system concerning the use of the eye movement in an online environment using CNN was clearly presented. Outstandingly, it is important to emphasize that the work concerning the AAC system yielded great results, as presented by the fact that the system yielded 99.9 percent accuracy concerning face recognition for a human, typing speed with a precision of 93 percent at a rate of 7.3 seconds/character, and finally yielded 100 percent accuracy concerning the recognition of characters. Quite evidence shows that the AAC technology yield was successful. In a similar vein, Aldaqre et al. (2024) examined various modalities of gaze pointing for AAC and proved that it significantly improves the accessibility of communication for people suffering

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2386-2392
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0320

from disability. They analyzed the dwell time selection method, the blinking selection mechanism, and smooth pursuit tracking for different kinds of users. The findings of the research indicated that the preference for a given method depends on the motor skills level of individual users as well as on the severity of disability. The results of many research studies have proved that the amalgamation of large language models with eye gaze pointing is a very effective option for handling complex users. Cai et al., in their study (2024), implemented a completely LLM-based AAC solution for ALS patients, providing a picture of how LLM incorporation, with a gaze type approach, enables faster speed in text typing and consequently enhances communication capabilities. In the research conducted by Dube and Wilkinson on "The role of eye tracking technology in the study of the patterns of visual engagement which are presented by persons with developmental disabilities who use Augmentative and Alternative Communication systems," the effectiveness of the usage of eye-tracking technology as an effective tool in understanding and researching the mechanisms by which a person with a developmental disability, such as ASD, CP, or ID, focuses his or her attention, which is very inadequately represented in most assessment tools, is made evident through the use of modern technology, such as those found in eye-tracking, which have shown promise in researching the impact of AAC, specifically its impact on independence, as well as its impact on communication. Research is being conducted to show the impact of AAC, to be more specific, its impact on independence, as well as its impact on communication. A mega-review research paper by Crowe et al. was con- ducted in 2023 to explore 84 studies published from 2000 to 2020. It followed critical appraisal guidelines defined by a tool called AMSTAR2. This is a major meta-analysis paper that established a considerable quantity of evidence that indicates that current AAC interventions, like PECS and SGD, play a major role in affecting successful outcomes in various disabilities. According to Crowe et al., they established that a person with a disability can successfully communicate with other people through AAC, thus successfully expressing his needs by communicating effectively with the help of AAC. This relates, among other things, to the fact that an artificial intelligent webcam eye tracking system, which has the potential for predicting screen gaze coordinates in real-time and without any other hardware requirement, as shown in Figueroa (2022), is yet another factor for the cost-effectiveness of using technology in AAC. Eye tracking technology has over time been known for its hardware requirement of thousands of dollars. The concept of computer vision utilized in the implementation of the proposed gaze tracking system is a multi-stage process that principally involves feature extraction related to the face of a human being. First, video frame capture techniques have been utilized, with this captured video being processed into grayscale form. In addition, through the use of front facial detectors with 68-point predictor integration, face detection has taken place, which principally involves ocular-related feature coordinates, followed by binarization techniques for face-related features, mainly considering various points of interest such as pupils and sclera. Subsequently, the gaze direction may be quantized using a ratio-based approach. Here, comparison of the pupil center coordinate with regard to boundaries established within the horizontal and vertical dimensions of the eye socket will lead to normalization between 0.0 and 1.0. In essence, this may be mapped to correspond with a specific state, such as left, right, and center, while using predetermined thresholds. Furthermore, there is scope for the feedback mechanism, which may be facilitated in real-time by using OpenCV and superimposing a green-colored crosshair over the original image [11-15].

3. Methodology

Figure 1 It depicts the multilingual gaze-based AAC system that uses a camera to acquire facial images at 30 FPS. Additionally, there is region of interest detection of the eye, face, and head using robust preprocessing for demographic accuracy. For the gaze prediction, the approach uses a CNN classifier named FullModel with multimodal inputs as well as uses temporal smoothing and maintains histories for language- specific buffers. Multilingual corpora, tokenizers (mBERT, XLM-R), and the n-gram models enable efficient handling of different scripts. Personalized word recommendations utilize an

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2386-2392
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0320

LSTM as well as transformer models according to incremental learning.

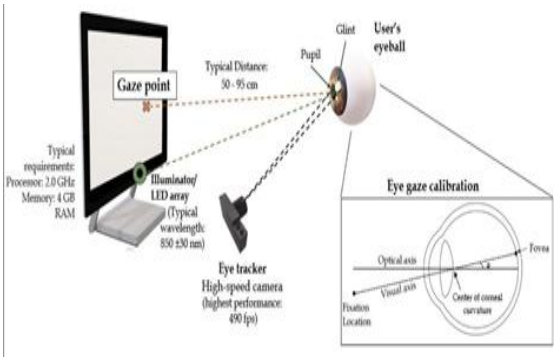

Figure 1 System Framework for Multilingual Based AAC

3.1. User Image Capture

It captures face images at 30fps with real-time detection and offers language options through an adaptive user interface [16-19].

3.2. Detection of Regions of Interest

The regions like faces are located by a cascade and a neural net, while pre-processing is done by" histogram equalization." This enhances contrast so that various shades of skin tones are taken care of. The head angle θ is useful for head estimation, as discussed above for gaze mapping

$$h(v) = \text{round}\,\frac{cdf(v) - cdf_{min}}{M \times N - cdf_{min}} \times (L-1)$$

3.3. Prediction CNN or Gaze Estimation

The FullModel CNN receives multimodal input data and estimates gaze coordinates (x^, y^) with smoothing from previous frames.

$$(\hat{x}, \hat{y}) = f_\theta(I_{\text{eye-left}}, I_{\text{eye-right}}, I_{\text{face}}, H_{\text{pos}}) \quad ; \quad \hat{x}_t = \frac{\sum_{i=1}^{n} w_i x_{i,t}}{\sum_{i=1}^{n} w_i}$$

3.4. History or Context Manager

Temporal buffers store past predictions, and a weighting mechanism prioritizes recent ones

$$\mu_t = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \quad ; \quad R[i,j] = R[i,j] + 1$$

3.5. Multilingual Corpus and Tokenization

The system incorporates the use of statistical-based n- gram systems in combination with other models such as the transformer-based architectures of mBERT and XLM-R.

3.6. Word or Sentence Recommendation

The model uses LSTM techniques, as well as transformer attention, in order to come up with the recommendations. The model can perform context-aware and personalized recommendations, as well as incremental learning [20 - 23].

3.7. Word or Sentence Recommendation

These recommendations involve the use of LSTM network- based recommendations and the use of transformer-based models with attention mechanisms. These models have the capacities for context-based and personalized recommendations and incremental learning Shown in Figure 2.

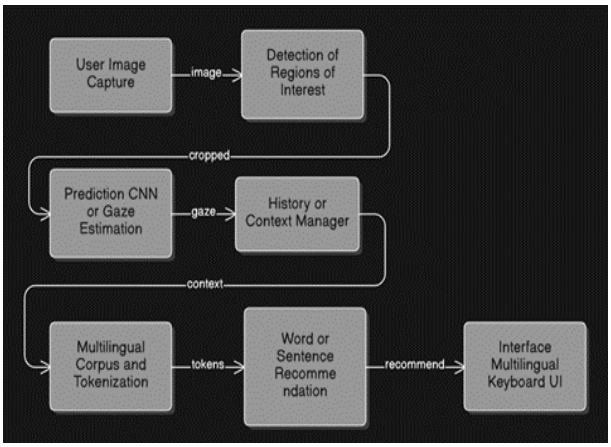

Figure 2 System Architecture

4. Experimental Assessment

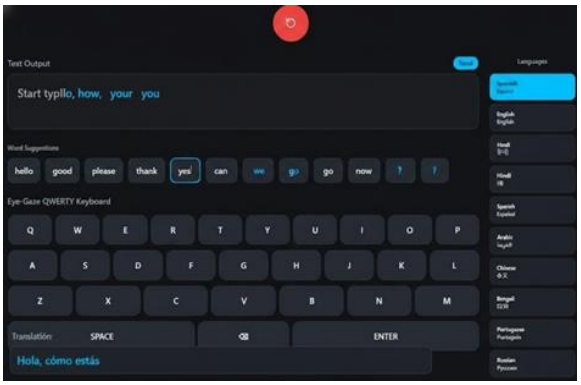

Figure 3 User Interface of Multilingual AAC System

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2386-2392
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0320

The efficiency of the proposed model of SEENS could be determined by using different standardized parameters such as classification, accuracy, precision, recall, and F one score. In this context, it could be asserted that the proposed model of SEENS has been compared with different state-of-the-art models such as Cell GAN and Res Skip CNN RF, which is briefed in Tables 1 to 4. As per the results, it could be evidenced that the proposed model of SEENS is found to be highly efficient and stable with a practically same accuracy of almost ninety-two, i.e., under different training epochs or during different experimental scenarios. Although Res Skip CNN RF model is achieving higher accuracy of almost ninety-five, it could be asserted that the proposed model of SEENS is highly efficient and could be implemented in a real-time scenario involving a webcam due to less computational resources. In addition, it could be asserted that the proposed model of SEENS has achieved higher precision or almost ninety with less gaze selection error and with a high recall rate of eighty- six during a typing scenario Shown in Figure 3 - 4.

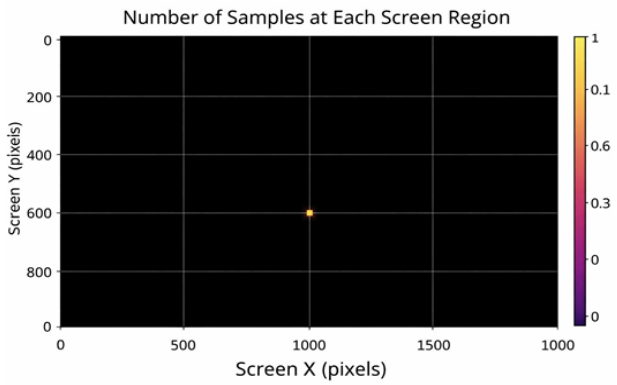

Figure 4 Number of Sample Regions

Table 1 Comparison of Accuracy (%)

Epochs	SEENS	Cell-GAN	Res Skip CNN-RF
10	92.1	93.2	95.2
20	91.1	93.1	95.0
30	92.4	93.2	94.9
50	92.3	93.1	95.0
65	92.0	93.1	95.1

However, standard values of metrics, which are usually employed for evaluation concerning the measure of efficiency of proposed models of SEENS, include accuracy, precision, recall, and, of course, F1-score. From the experimental results, it can be stated that SEENS has a certain degree of stability related to its value of accuracy because, even in conditions of employing different values of training epochs, its value remains unchanged, reaching an accuracy level of approximately 92% [24 - 27].

Table 2 Comparison of Precision (%)

Epochs	SEENS	Cell-GAN	Res Skip CNN-RF
10	90.0	91.0	93.0
20	90.1	91.2	93.1
30	90.2	91.2	93.2
50	90.0	91.3	93.0
65	90.1	90.9	93.1

Table 3 Comparison of Recall (%)

Epochs	SEENS	Cell-GAN	Res Skip CNN-RF
10	86.2	88.1	89.0
20	85.9	88.0	89.3
30	86.0	88.2	89.3
50	86.0	88.0	89.2
65	86.0	87.9	89.2

Table 4 Comprehensive Comparison of Existing and Proposed Methodologies (%)

Metric	SEENS	Cell-GAN	Res Skip CNN-RF
Accuracy	92	93	95
Precision	90	91	93
Recall	86	88	89
F1-Score	81	82	85

It is also important to note that its peak value is close to 95%, which is characterized by higher computational complexity, being enough for employing SEENS using a real webcam. Besides, its value of precision is estimated as approximately 90%, but only a small amount of single gaze is available and future work.

Conclusion

- In this research, a powerful yet cost-effective system has been specified for Augmentative and Alternative Communication (AAC)

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2386-2392
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0320

systems, as it allows for eye and head tracking through CNN-based techniques supported by standard webcams. The proposed system also optimizes the typing rate while avoiding the Midas Touch issue through multi-fold triggers like blinking and wink identification. Additionally, providing multi-lingual options and adaptable key sizes improves the usability of the specified system. In future work, the focus will be on lowering the cognitive load, improving privacy safeguards, and thus providing a high-performance and user-centric communication lifeline for individuals with significant motor difficulties.

- While the current framework serves to prove the efficacy of the webcam-based AAC systems, there are indeed some fruitful paths that can be pursued with regards to future research and development of this technology in the near future. Its development would focus on ensuring environmental robustness, artificial intelligence, as well as multi-modal biological signals.
- Environmental and Hardware Robustness These limitations, which have been affecting RGB-based eye-tracking systems, in particular, are related to environmental as well as hardware robustness. For instance, in this new version of proposed eye-tracking systems, it has been proposed that the system will be able to accommodate the algorithms like" Adaptive Gain Control" as well as" Contrast Enhancement" in later versions. Notably, once this new feature has been added to the proposed eye-tracking system, it would be feasible to utilize this particular system even in low environmental conditions. In this proposed version of the system, it has been proposed that the system will be able to accommodate even the" Head Pose Compensation" feature so that the movement of the head can also be detected. For instance, the proposed system will be helpful to those individuals who are not able to move their head due to tremors.
- Emotion-Aware Predictive Modeling The current DistilGPT-2 engine predicts words based on linguistic probability. Future development will test the integration of Affective Computing, whereby the system analyzes the user's face expressions alongside their gaze. If this can detect a user's emotional state, then the Transformer model could highlight certain vocabularies or" Quick-Action" phrases related to their immediate needs (e.g., medical alert or emotional responses) and further reduce the cognitive load.
- Multi-Modal Interaction and BCI Integration in order to advance towards a more integrated tool, we would like to explore the possibility of integrating gaze-tracking methods with other low-cost devices such as Electroencephalogram (EEG) or Electromyography (EMG) headbands. This would enable the potential for a" hybrid click" based on" intent to blink" or brainwaves, effectively eliminating delays based on dwell time altogether. This will enable the system to cater to those affected by degeneratives in more advanced stages, where muscle control may be impaired.

References

[1]. H. Wang et al., "Quantifying the impacts of posture changes on office worker productivity," BMC Public Health, vol. 23, 2023.

[2]. R. A. da Silva and A. C. Paschoarelli Veiga, "Algorithm for decoding visual gestures for an assistive virtual keyboard," IEEE Latin America Transactions, vol. 18, pp. 1909–1915, 2020.

[3]. A. McNicholl et al., "The impact of assistive technol- ogy use for students with disabilities: A systematic review," Disability and Rehabilitation: Assistive Technology, 2021.

[4]. F. L. Silva and A. R. C. Serra, "Assistive technology: Augmentative and alternative communication resources," Re- vista Tempos e Espac¸os em Educac¸a˜o, 2020.

[5]. L. F. B. Loja et al., "An evolutionary virtual keyboard for alternative communication systems," 2021.

[6]. A. C. A. Montenegro et al., "Use of a robust

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2386-2392
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0320

alternative communication system in autism spectrum disorder: A case study," Revista CEFAC, 2022.
[7]. C. M. Togashi and C. C. de Figueiredo Walter, "Con- tributions of alternative communication use in the school inclusion process," 2021.
[8]. R. Bonotto et al., "Learning opportunities supported by augmentative and alternative communication during the COVID-19 pandemic," Ibero-American Journal of Educational Studies, 2021.
[9]. A. Anandika et al., "Hand gesture control of a virtual keyboard using neural networks," 2023.
[10]. D. Freitas, S. Rodrigues, and J. Ribeiro, "Computer access interfaces for people with motor impairments: A state- of-the-art review," 2021.
[11]. M. Nazar et al., "A systematic review of hu- man–computer interaction and explainable AI in healthcare," IEEE Access, 2021.
[12]. J. Hori et al., "Development of a communication support device controlled by eye movements and voluntary eye blinks," in Proc. 26th Annual Int. Conf. IEEE EMBS, 2004.
[13]. J. O. Wobbrock et al., "Not typing but writing: Eye-based text entry using letter-like gestures," in Proc. CHI, 2008.
[14]. A. Bulling, D. Roggen, and G. Tro¨ster, "It's in your eyes: Toward context awareness and mobile HCI using wear- able EOG goggles," 2011.
[15]. W. Tangsuksant et al., "Directional eye movement detection system for virtual keyboard control," 2018.
[16]. H. Cecotti, Y. K. Meena, and G. Prasad, "A multi- modal virtual keyboard using eye tracking and hand gesture detection," 2016.
[17]. S. Tantisatirapong and M. Phothisonothai, "Design of a user-friendly virtual Thai keyboard based on eye tracking," 2021.
[18]. M. I. Rusydi et al., "Adaptive symmetrical virtual keyboard based on EOG signals," 2019.
[19]. A. Z. Attiah and E. F. Khairullah, "Eye-blink detection system for virtual keyboard interaction," 2021.
[20]. H. Drewes, "Eye gaze tracking for human–computer interaction," 2010.
[21]. K. H. Holmqvist et al., "Eye tracking: Empirical foun- dations for a minimal reporting guideline," Behavior Research Methods, 2022.
[22]. D. E. King, "Max-margin object detection," 2015.
[23]. D. da Silva Lima, Evaluation of Child Visual Function Using an Automated Video-Based Eye Tracking Solution, Ph.D. dissertation, University of Sa˜o Paulo, 2021.
[24]. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[25]. J. Maher, The Future Was Here: The Commodore Amiga. Cambridge, MA, USA: MIT Press, 2012.
[26]. N. Garay-Vitoria and J. Abascal, "Text prediction sys- tems: A survey," Universal Access in the Information Society, 2006.
[27]. L. Florea et al., "Can your eyes tell me how you think? Gaze-directed estimation of mental activity," 2013.