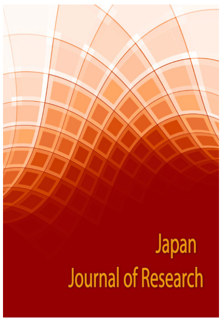
Japan Journal of Research

Original Article



Prediction for Occurrence of Character Identification Difficulty during Web Browsing Based on Gaze Data

Matsushita Yutaka and Aoki Shusei

Department of Media Informatics, Kanazawa Institute of Technology, Japan

Correspondence

Matsushita Yutaka

Department of Media Informatics, Kanazawa Institute of Technology, Japan

Abstract

This study develops a system to predict, with high precision and in real-time, the occurrence of difficulty in character identification during web browsing, based on gaze data. Specifically, the system leverages fixation duration, which evolves incrementally, and employs a reinforcement learning algorithm based on SARSA, to evaluate the occurrence of the difficulty at each step. Since fixation durations caused by character identification difficulty are not necessarily longer than those resulting from other factors, establishing a reliable threshold for character magnification is difficult. Nevertheless, the system must refrain from magnifying characters when users do not feel them difficult to identify. Therefore, this study introduces saccadic velocity and amplitude as two external parameters, categorizes them into distinct groups, and calculates the Q-value for each category pair, thereby enabling a precise determination of magnification thresholds. Furthermore, a method for assigning rewards and penalties to the agent is examined.

- Received Date: 10 Aug 2025
- Accepted Date: 22 Aug 2025
- Publication Date: 25 Aug 2025

Keywords

Reinforcement learning, SARSA, Character identification, Fixation duration, Saccadic amplitude, Saccadic velocity

Introduction

Although the utilization of ICT has increased among older adults, many individuals face significant challenges in web browsing because of deteriorating eyesight. To promote an inclusive society, it is essential to facilitate seamless web browsing experience for individuals with mobility impairments. Eye gaze-controlled web browsers [1, 2] have been developed to assist users with such impairments. An active browsing system is desired to both predict and preemptively address potential issues encountered by users. As shown in Figure 1, it is convenient for a browsing system to automatically enlarge text when users fixate on characters they find difficult to identify during web browsing. The development of such a browsing system requires the avoidance of two types of errors: the first occurs when characters are enlarged despite a user not experiencing difficulty in identification (**Error 1**), and the second occurs when characters are not enlarged, even though the user experiences difficulty in identification (**Error 2**). However, accurately predicting the onset of difficulty in character identification (which ought to be essential to decrease the two types of errors) is challenging because gaze fixation may result from various factors other than difficulty in character identification, such as issues with sentence comprehension. The fixation duration associated with difficulty in character identification is not necessarily longer than that associated with other causes (Figure 2). Consequently, it is reasonable to implement an AI system that learns the browsing patterns of individuals and enlarges characters based on these learned behaviors.

Reinforcement learning [3] is the most suitable machine learning algorithm for this task. Although complete training data are not available, the system can validate the appropriateness of character enlargement by querying users to confirm whether their gaze was directed on the character owing to difficulty in identifying it. This enables the system to learn the eye movement patterns of users who find it challenging to identify a character in a trial-and-error manner (reinforcement learning) to obtain an optimal method for character magnification. Notably, in contrast to usual reinforcement learning, the subsequent restriction is imposed on the prediction system because users request the system to enlarge characters by directing their gaze on the characters for a specified duration: "the system must enlarge the characters on all fixation points whose fixation durations are longer than a threshold." Given the critical role of this threshold, this study employs SARSA learning [4] (rather than *Q*-learning [5]) as an algorithm for reinforcement learning. SARSA is particularly well-suited to this task because it allows the system to update its action (character enlargement or non-enlargement) based on actual user actions (for instance, refer [6] for a comparison between SARSA and *Q*-learning), thereby increasing the likelihood of accurately identifying an appropriate threshold.

Citation: Yutaka M, Shusei A. Prediction for Occurrence of Character Identification Difficulty during Web Browsing Based on Gaze Data. Japan J Res. 2025;6(10):158.

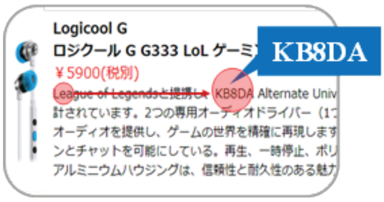


Figure 1: *Browsing system automatically enlarging characters*

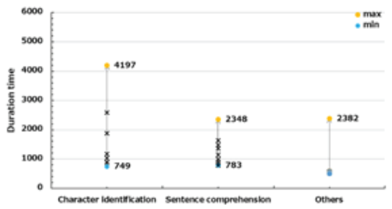


Figure 2: *Fixation durations according to causes.*

The aim of this study is to develop a system that can predict rapidly and accurately the occurrence of difficulty in character identification from users' gaze fixations through reinforcement learning. In gaze data analysis [7, 8] via reinforcement learning, the state is typically defined as the duration of each fixation point. However, this approach lacks rapid prediction because it relies on fixation durations that have already passed. To address this limitation, this study proposes a novel approach in which the fixation duration at each fixation point increases sequentially and an agent (decision maker) determines whether difficulty in character identification occurs in each increment of the fixation duration. In other words, by treating each progressive time point of gaze fixation as a state of reinforcement learning, we enable the agent to enlarge a character just before the user has difficulty. Additionally, setting an upper limit of 3,000 ms on the fixation progression time for the enlargement of characters reduces the occurrence of Error 1. Figure 2 indicates that setting a uniform standard (fixation duration) for enlarging characters is nearly impossible. However, such a criterion is necessary because of the abovementioned restriction. To address this problem, this study introduces the saccadic velocities and amplitudes of users as external parameters, and provides criteria based on the combination of the categories of these two parameters. This approach is expected to improve accuracy because the criterion can be customized for each category. Moreover, this treatment aligns with psychological [9, 10] and ergonomic [11] research, which have reported that both fixation time and the saccadic velocity and amplitude vary when users encounter situations requiring judgment during information search or text reading.

Specifically, the appropriate number of categories for the two parameters and the method for providing rewards for the agent are explored. The reduction of the frequency of Error 1 is given priority over that of Error 2 based on the understanding that users find Error 1 more uncomfortable and are more likely to tolerate slight delays, given that this is a gaze-controlled magnification.

Related work

Extensive research has been conducted on gaze-controlled websites. Menges et al. [1] developed a web browser, GazeTheWeb, with direct gaze control for individuals with motor impairments. By understanding interface semantics, they mapped the functions of interaction elements with suitable gaze interaction, aiming to reduce the cognitive load associated with controlling an interface via eye gaze input. Furthermore, in comparison with the familiar (indirect) gaze-control browser, OptiKey [12], they demonstrated the superior performance of GazeTheWab in terms of response time and subjective usability evaluation. Kumar et al. [2] developed a secure and fast PIN entry system by effectively combining gaze and touch modalities.

Meanwhile, the development of active web browsing systems has progressed to provide convenient services by predicting user trends and potential future issues. Salem's [13] work on user interface optimization based on Genetic Programming is a pioneering example of such a system. Along with the rapid advancements in AI, research [14,15] on applying Neural Networks to eye-tracking technology is growing. However, reinforcement learning is particularly suitable for active web browsing. This is because the agent can explore and find optimal services for a user through trial-and-error, utilizing "partial" training data without requiring complete supervised signal. Deep *Q*-network, a combination of *Q*-learning and Neural Networks, has a significant advantage in handling large data sets [16] because of its ability to virtually create training data by incorporating neural networks. Therefore, deep *Q*-networks are primarily applied to the development of recommender and web navigation systems [17-19]. For instance, in the development of recommender systems, the user access logs over a constant period can serve as an effective (partial) training data because they are collected naturally and reflect the user's behavior. In contrast, (not deep) *Q*-learning or SARSA is typically used to analyze personal gaze behavior [7, 8, 20], such as prediction of the next fixation point during reading (although this differs from the context of active web browsing). In general, tracking a user's gaze movement is more complex than collecting access logs. Additionally, training data generated virtually by Neural Networks may not accurately reflect user-specific properties, which is different from the recommendation systems.

This study adopts SARSA learning because gaze data are necessary to predict the occurrence of difficulty in character identification. However, the method of defining states is entirely different from previous studies [7, 8, 20]. Notably, the uniqueness of this study lies in the introduction of reinforcement learning and eye tracking technology to achieve "active web browsing.".

Method

Basic Concepts

In reinforcement learning, an agent (decision maker) chooses an action in each state and reaches the next state; a reward is received if the choice is successful. The multiplication of a reward by -1 means that a penalty is charged if the choice fails. The agent learns how to choose an action in each state such that an expected future return is maximal. This is mathematically formulated as follows. Let s_t, a_t, and r_t be a situation, action, and reward at time step t, respectively; sample values are written in lowercase, assume that capitalization, say S_t, A_t, and R_t, denotes random variables. Let $Q(s_t,a_t)$ denote an expected (future) return when the agent chose a_t in s_t, called an **action value function** or ***Q*-value**, which is written in the form.

$$Q(s_t,a_t)=\max_{a_t\in\mathcal{A}} E\left\{R_t+\gamma R_{t+1}+\gamma^2 R_{t+2}+\cdots|S_t=s_t,A_t=a_t\right\}$$

where $\mathcal{A}$ is the set of all actions and γ is a discount rate. Given that the environment model is completely known, it is possible to calculate the above expectation directly. Otherwise, the calculation must be conducted based on the sample approximation. SARSA learning is valid when sampling data are obtained sequentially, and one wishes to update the Q-value using only new (collected) data at one time step. SARSA learning is a type of Temporal Difference learning, and its update formula is expressed as follows:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\delta_t,\ \delta_t = r_t + \gamma Q(s_{t+1},a_{t+1}) - Q(s_t,a_t) \quad (1)$$

The arrow of Eq. (1) implies that the right-hand updates the left-hand side, α is a learning rate, and δ_t, called a TD error, is the difference between Q-values estimated at time steps t+1 and t. Note that an actual action is used in the update in SARSA. Finally, the agent can select the action in each state based on

$$a_{t,\max} = \underset{a_t \in \mathcal{A}}{\operatorname{argmax}}\, Q(s_t,a_t) \quad (2)$$

States

An agent is a user's personal computer, and the actions taken by the agent include enlargement and non-enlargement. A key requirement for the prediction system in this study is that, when a user feels difficulty in identifying a character and focuses their gaze on it, the agent must enlarge the character as quickly as possible. To achieve this, a state is defined as the elapsed time when fixation duration evolves step by step. If the time increment is Δd, the state (gaze fixation time) s_t at time step t is expressed by .

$$s_t = F_{\min} + \Delta d \cdot (t-1),$$

where F_{min} is the minimum fixation time (500 ms) for all gaze points. The formula implies that the time steps and states correspond one-to-one. Specifically, the time interval from 500 to 3,000 ms was divided into 100 ms intervals so that 25 discrete fixation times were yielded. This enables the agent to judge whether the enlargement of characters is conducted at each level of fixation time and shorten the response time. Table 1 lists the time steps, states, and an example set of actions that can be chosen for the states using Eq. (2).

Each fixation point corresponds to a game (episode) of Go or Chess, and the choice of an action in each state corresponds to a move of the game. However, the prediction system developed in this study is significantly different from the usual reinforcement learning in that a restriction (i.e., once the character enlargement is determined at a specific level of the fixation duration, then it is conducted for all fixation points whose durations exceed the level) is imposed because the agent is requested to enlarge the character following the user's gaze fixation. This restriction does not imply that reputation for prediction is acquired by simply finding a set of correct actions at each fixation point. It is unlikely that, beyond a certain level, the Q-values of enlargement are always greater than those of non-enlargement. Table 1 illustrates the difficulty in setting a threshold for enlarging the characters. Setting the threshold at a low (2,700) and a high level (2,900 ms) increases the frequency of Errors 1 and 2, respectively. Nevertheless, a uniform threshold for determining whether to enlarge a character must be derived across all the gaze points, even if the results (correct actions) of several gaze points are sacrificed. To minimize the number of sacrificed results as small as possible, it is suitable to introduce the saccadic velocity and amplitude as external conditions and derive such a threshold within each group of fixation points with the same metric values. Specifically, let sv_k and sa_k be a saccadic velocity and amplitude just before a fixation point k. First, the ranges (maximal minus minimum values) for the velocity and amplitude are divided into n categories and the respective category numbers to which sv_k and sa_k belong are searched. Second, a Q table is constructed for each pair of category numbers. For example, if sv_k and sa_k belong to the i-th and j-th categories ($1 \le i, j \le n$), respectively, then the Q values indexed by (i, j) are obtained, denoted $Q_{(i,j)}$. Finally, using $Q_{(i,j)}$, we find the minimum of those s' such that a_t = enlargement whenever $s_t > s'$. The threshold Fix(x_i) to be derived is defined as the minimum, which specifies the rule

$$s_t \ge \mathrm{Fix}_{(i,j)} \Rightarrow \text{enlargement};\ s_t < \mathrm{Fix}_{(i,j)} \Rightarrow \text{non-enlargement}. \quad (3)$$

Table 1: *Time steps, states, and chosen actions*

t-step	state	Action
1	s_t = 500	Non-enlarge
2	s_t = 600	Non-enlarge
3	s_t = 700	Non-enlarge
⋮	⋮	⋮
23	s_t = 2700	Enlarge
24	s_t = 2800	Non-enlarge
25	s_t = 2900	Enlarge

Rewards and penalties

To reduce the occurrence frequency of Errors 1 and 2, this subsection elaborates on a method for assessing rewards and penalties. Henceforth, let P (positive) and N (negative) correspond to enlargement and non-enlargement; both have a one-to-one correspondence with difficulty and non-difficulty in identifying characters. **Precision** and **recall** are performance metrics for prediction models, which are defined by the formulas.

$$\frac{TP}{TP+FP} \text{ and } \frac{TP}{TP+FN},$$

where TP denotes the number of choices in which enlargement is correct, and FP and FN denote the number of choices in which enlargement and non-enlargement are incorrect, respectively. Since the decreases in FP and FN induce those of Errors 1 and 2 by the formulas, it follows that the proper evaluation of rewards and penalties is such one that it causes the decreases in FP and FN.

This consideration prompts the formulation of the following policy on the evaluation:

1. To decrease FP, the agent is prevented from enlarging the character for short fixation durations as much as possible. For short fixation durations, only a very small reward is given to the agent even if the enlargement is correct, and only a small penalty is imposed if the non-enlargement is incorrect.
2. To decrease FN, the agent is forced to enlarge the characters when the fixation duration is long. For this, if the agent does not enlarge characters in the super-long-time domain and the choice is incorrect, then an extremely large penalty is imposed on them.
3. We assume that rewards for enlargement monotonously increase, and penalties monotonously decrease as the fixation time progresses. The former assumption is to make the agent's choice consistent with the compulsory enlargement of more than or equal to 3,000 ms. The latter is based on the natural expectation that a long

non-enlargement accompanied by failure makes users uncomfortable.

Precision and recall can be used as assessment metrics for the decreases of Errors 1 and 2. Specifically, let R=100 be a basic reward; a **magnification factor** $c_t(1 \le t \le 25)$ is determined for each time step t and each reward and loss are calculated by multiplying R by this coefficient:

$r_t = c_t \cdot R$ where $c_t > 0$ for rewards and $c_t <$ for penalties.

Verification of model

Experiment

The stimulus was a virtual website featuring computer peripherals. The website contained images, names, and detailed descriptions of 15 products. Approximately 200 Japanese characters were used for each description. The layout was such that the products were listed vertically on one page and no page transitions exited. The font used was "MSP Gothic" with a size of 16 px, which was used in the web site of Yahoo Japan.

The subject was a graduate student (male, 23 years old) in Kanazawa Institute of Technology. The participant was requested to browse the stimulus site freely, presented on a 14-inch laptop PC (Panasonic CF-LF), and click the mouse when difficulty in character identification arose. During web browsing, gaze movements of the participant were recorded using an eye tracker (Tobii Pro Spark; 60 Hz). The saccadic amplitude was defined as the distance between each pair of consecutive fixation points and the saccadic velocity at a certain point was calculated by dividing the saccadic amplitude at the temporal point by the difference between the two sampling times. Thus, all fixation durations, saccadic velocities, and amplitude were obtained during this browsing. As a result, the total number of fixations (duration greater than or equal to 500 ms) was 455, in which the number of fixations when the participant felt difficult to identify a character and fixed his eyes on it was 25.

Model 1 (without external conditions)

Rewards and penalties were assigned to each state in the prediction model into which saccadic velocities and amplitudes were not introduced. Table 2 lists the magnification factor values for *TP*, *TN*, *FP*, and *FN* in each state. The table shows that for

Table 2: *Magnification factor values*

		Reward		Penalty	
t-step	state	TP	TN	FP	FN
1-10	$500 \le s_t \le 1400$	0.25	1.50	-7.00	-0.25
11-22	$1500 \le s_t \le 2600$	0.25	1.00	-6.00	-0.50
23,24	$2700 \le s_t \le 2800$	3.00	0.25	-6.00	-10.00
25	$s_t = 2900$	5.00	0.25	-6.00	-15.00

TP and *FN*, the absolute values were small in the range of short fixation durations (Policy 1), while for *FN*, the absolute value was extremely high at S_t = 2,900 (Policy 2).

Using these values and Eq. (1), Q-values were calculated. The number of iterations was 10,000; however, convergence was achieved after about 6,000 iterations. The threshold of Eq. (3) was 2,700 ms. This model had

Precision = 0.94 and Recall = 0.60. (4)

Thus, the occurrence frequency of Error 1 appeared to be comparatively low, whereas that of Error 2 was high. This result is attributed to our attempt to decrease the frequency of Error 1 as much as possible. However, from a practical point of view, it is uncertain whether the frequency level is acceptable.

Model 2 (with external conditions)

Using the rewards and penalties listed in Table 2, a prediction model was constructed by introducing saccadic velocities and amplitudes. The maximum and minimum of saccadic velocities were 9.029 and 0.013 px/ms, and those of saccadic amplitudes were 775.7 and 4.123 px. The ranges of these metrics were divided into 9 equal intervals to create 9×9 categories. Figure 3 shows the number of fixations in each entry in the 9×9 category matrix when the subject felt difficulty in identifying a character and fixed his eyes on it, where each numeral in parentheses indicates

Figure 3: Number of fixations in each entry

i \ j	1	2	3	4	5	6	7	8	9
1	3 (28)	0 (6)	0 (4)	0 (1)	0 (4)		1 (1)	0 (3)	0 (3)
2	3 (22)	2 (19)	0 (7)					0 (1)	0 (2)
3		2 (26)	0 (8)	1 (7)	0 (9)				
4			2 (30)	1 (4)	0 (2)	0 (11)	0 (4)		
5			0 (1)	2 (36)	0 (2)		0 (11)		
6				0 (3)	1 (33)	1 (6)		0 (9)	
7						3 (34)	0 (6)	0 (9)	0 (1)
8							2 (27)	0 (20)	0 (4)
9								0 (10)	1 (41)

Figure 4: Threshold values in each entry.

i \ j	1	2	3	4	5	6	7	8	9
1	2000						500		
2	1700	1700							
3		2500		1200					
4			2200	1100					
5				2500					
6					2100	2200			
7						2700			
8									
9									2700

Figure 5: Precision in each entry

i \ j	1	2	3	4	5	6	7	8	9
1	1						1		
2	1	1							
3		1		1					
4			1	1					
5				1					
6					1	1			
7						0.75			
8							—		
9									1

Figure 6: Recall in each entry.

i \ j	1	2	3	4	5	6	7	8	9
1	0.333						1		
2	0.667	0.5							
3		0.5		1					
4			0.5	1					
5				1					
6					1	1			
7						1			
8							0		
9									1

the total number of fixations. The frequency of fixations caused by difficulty in character identification was extremely low in each entry. Fixations were concentrated around the diagonal of the category matrix, suggesting a high correlation between velocity and amplitude (correlation coefficient = 0.76). Hence, it appears that the use of either velocity or amplitude data was sufficient for prediction. However, since samples exist except for the diagonal and sample data will be large in the future, it is rational to use both data.

Each Q-value was calculated in the (i, j)-th entry of the category matrix. Figures 4, 5, and 6 show the threshold $\mathrm{Fix}_{(i,j)}$ and the precision and recall scores in the (i, j)-th entry, respectively. As shown in Figure 4, $\mathrm{Fix}_{(i,j)}$ tended to be somewhat larger as the saccadic velocity and amplitude increased, although further verification is necessary because the number of samples was small and there was only one subject. Figures 5 and 6 suggest that the Q-values with indices improved precision so that the values were 1.0 in almost all entries but did not improve recall at such a level; indeed, the values were 1.0 in about half entries. To compare these results with those of the model in the previous subsection, the comprehensive precision and recall were calculated by weighing the values of each entry by the sample numbers:

$$\text{Precision} = 0.96 \text{ and Recall} = 0.68. \quad (5)$$

Comparison of Eqs. (4) with (5) seems to indicate that there might be little improvement, even in terms of precision. This was primarily due to the small sample size. Indeed, there was only one error, which gave rise to a poor precision score (0.75) for the (7, 6)-th entry. Changing the evaluation of rewards and penalties through all samples into that through samples of each entry would almost certainly make each precision of the entry 1.0; the recall scores will also be expected to improve. Moreover, it is expected that the larger the sampling number, the greater is the effect of these fine thresholds on the reduction of Errors 1 and 2.

Conclusions

This study developed a system to predict the occurrence of difficulty in character identification quickly and correctly from users' gaze data during web browsing, aiming to realize an active web equipped with automatic magnification of characters. To enable rapid prediction, a state was defined as the elapsed time when the fixation duration enveloped incrementally. An algorithm to assess the occurrence of difficulty at each time step was created based on SARSA. Moreover, to minimize two types of prediction errors, the Q-values indexed by categorized saccadic velocities and amplitudes was calculated and a novel method was devised to evaluate rewards and penalties. An evaluation experiment demonstrated that this prediction method could be effective in reducing the two types of errors if the amount of sample data were increased. Future directions for this study include the following: First, the model must be validated by increasing the number of participants. Second, the development of predictions based on DQN learning is necessary to construct an active website capable of automatically magnifying characters, which would require handling a substantially large sample data.

Conflicts of Interest

No potential conflict of interest was reported by the article.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers 21K12096 and 23K26179. The authors would like to express their gratitude to Fukuyo Tsuyoki (2024 undergraduate), a student of the first author, for his assistance in collecting gaze data and making predictions.

References

1. Menges R, Kumar C, Staab S. Improving user experience of eye tracking-based interaction: introspecting and adapting interfaces. ACM Trans Comput Hum Interact. 2019;26(6):37.
2. Kumar C, Kabari D, Menges R, MacKenzie IS, Staab S. TouchGazePath: multimodal interaction with touch and gaze path for secure yet efficient PIN entry. In: Proceedings of the International Conference on Multimodal Interaction (ICMI '19); October 2019; Suzhou, China. p. 329-338.
3. Sutton RS, Barto AG. Reinforcement Learning. 4th ed. Cambridge, MA: MIT Press; 2018.
4. Rummery GA, Niranjan M. Online Q-learning using connectionist systems. Cambridge University; 1994:23-86.
5. Watkins CHCJ. Learning from delayed rewards. Robot Auton Syst. 1989;15(4):233-235.
6. Xu Z, Cao L, Chen X, Li C, Zhang Y, Lai J. Deep reinforcement learning with Sarsa and Q-learning: a hybrid approach. IEICE Trans Inf Syst. 2018;E101(D9):1241-1251.
7. Nagarajan H, Inakollu VS, Vancha P, AJ. Detection of reading impairment from eye-gaze behaviour using reinforcement learning. Procedia Comput Sci. 2023;218:2734-2743.
8. Jiang M, Boix X, Roig G, Xu J, Van Gool L, Zhao Q. Learning to predict sequences of human visual fixations. IEEE Trans Neural Netw Learn Syst. 2016;27(5):1241-1252. doi:10.1109/TNNLS.2015.2496306
9. Rayner K. Eye movements and information processing: 20 years of research. Psychol Bull. 1998;124(3):372-422. doi:10.1037/0033-2909.124.3.372
10. Rayner K, Morris RK, Pollatsek A. Eye movement guidance in reading: the role of parafoveal letter and space information. J Exp Psychol Hum Percept Perform. 1990;16(2):268-281.
11. Goldberg JH, Stimson MJ, Lewenstein M, Scott N, Wichansky AM. Eye tracking in web search tasks: design implications. In: Proceedings of the 2002 Symposium on Eye Tracking Research & Applications (ETRA); 2002; New Orleans, LA. p. 51-58. doi:10.1145/507072.507082
12. Sweetland J. Optikey: Type, Click, Speak. GitHub. Published 2016. Accessed August 19, 2025. https://github.com/OptiKey/OptiKey
13. Salem P. User interface optimization using genetic programming with an application to landing pages. Proc ACM Hum-Comput Interact. 2017;1(1):12.
14. Severitt BR, Lenhart P, Hosp BW, Castner NJ, Wahl S. Communication breakdown: gaze-based prediction of system error for AI-assisted robotic arm simulated in VR. In: Proceedings of the 2024 Symposium on Eye Tracking Research & Applications (ETRA). 2024;5:1-7. doi:10.1145/3649902.3653339
15. Chugh S, Ye J, Fu Y, Eizenman M. CSA-CNN: a contrastive self-attention neural network for pupil segmentation in eye gaze tracking. In: Proceedings of the 2024 Symposium on Eye Tracking Research & Applications (ETRA). 2024;6:1-7. doi:10.1145/3649902.3653351
16. Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. Nature. 2015;518(7540):529-533.
17. Minoda K, Auchi H, Kawagashira N, Ishikawa N. Deep reinforcement learning for recommender system of users' behavior on website (in Japanese). In: Proceedings of the 33rd Annual Conference of the Japanese Society for Artificial Intelligence;

2019. 4O2-J-2-01:1-4.

18. Lapan M. Chapter 16: Web navigation. In: Deep Reinforcement Learning Hands-On: Apply Modern RL Methods to Practical Problems of Chatbots, Robotics, Discrete Optimization, Web Automation, and More. 2nd ed. Birmingham, UK: Packt Publishing; 2020:447-499.
19. Afsar MM, Crump T, Far B. Reinforcement learning-based recommender systems: a survey. ACM Comput Surv. 2022;55(7):145. doi:10.1145/3543846
20. Liu Y, Reichle ED, Gao DG. Using reinforcement learning to examine dynamic attention allocation during reading. Cogn Sci. 2013;37(8):1507-1540. doi:10.1111/cogs.12027