Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p26212-cd

Predicting Seafarer Control Transfer Intentions for Safe Human–Machine Collaboration in Maritime Autonomous Surface Ships

Wanning Li

State Key Laboratory of Maritime Technology and Safety, Wuhan 430063, China
School of Transportation and Logistics Engineering, Wuhan University of Technology, Wuhan 430063, China
National Engineering Research Center for Water Transport Safety, Wuhan 430063, China
E-mail: 378047@whut.edu.cn

Shiqi Fan*

State Key Laboratory of Maritime Technology and Safety, Wuhan 430063, China
School of Transportation and Logistics Engineering, Wuhan University of Technology, Wuhan 430063, China
National Engineering Research Center for Water Transport Safety, Wuhan 430063, China
E-mail: fanshiqi@whut.edu.cn

With the rapid development of autonomous ships, the role of seafarers onboard Maritime Autonomous Surface Ships (MASS) is in significant transformation. Among onboard operations in autonomous ships, the control transfer between seafarers and systems serves as a critical process to ensure MASS safety. However, research gaps remain concerning when and how seafarers takeover control from autonomous systems. This study investigates seafarers' control transfer intentions in hazardous navigation scenarios by integrating computer vision techniques with behavioral analysis in real-world sailing experiments. The experimental framework synchronously captures seafarer behavior videos and vessel status data onboard autonomous ship Yujiaotou 001. Computer vision methods are employed to extract key behavioral features, including body posture and hand movements. By integrating behavioral features with vessel trajectory data, this study identifies key indicators of seafarers' intention to shift and develops an intention prediction model for onboard seafarers during control transfer. The findings provide theoretical and empirical foundations for enhancing maritime education and training for MASS, as well as improving safety management in autonomous ship operations.

Keywords: MASS, human-machine collaboration, control transfer, takeover, handover

1. Introduction

With the advancement of digital and intelligent technologies, Maritime Autonomous Surface Ships (MASS) are driving the transformation of the global shipping industry (Li and Yuen, 2024). Guided by these technologies, MASS can reduce both labor costs (Munim et al., 2025) and physical workload, as well as lower emissions (Fan et al., 2020). The International Maritime Organization (IMO) classified MASS into four autonomous degrees in 2018 to clarify its technical boundaries and development paths. Before achieving full autonomy, ships of the D2 and D3 degrees still require human intervention, making takeover and handover processes integral to MASS operations. In such a human-machine collaboration framework, the ship Autonomous Navigation System (ANS), shore-based operators, and onboard seafarers collectively form a complex system. The ANS, serving as the intelligent terminal on MASS, is responsible for monitoring and controlling onboard equipment (Feng et al., 2022). Shore-based operators perform remote supervision and control through the Remote Operations Center (ROC), while experienced personnel onshore assume auxiliary roles. Given ANS is not yet fully capable of managing all complex navigational scenarios (Yan et al., 2021), ensuring safe and timely control transfer between systems and seafarers has become critical to mitigating navigational risks and preventing accidents in MASS operations.

Control transfer, including takeover and handover processes, is closely associated with navigational safety and requires a certain time budget (Wróbel et al., 2025) for operators to regain control. Although MASS technologies have developed fast, there is limited evidence to showcase when and how to transfer control between seafarers and autonomous systems. Their control transfer intention represents the decision making of seafarers given navigational risks and cognitive load. So understanding control transfer intention is the first step to investigate when and how seafarers intervene the autonomy to reduce predictive risks. A safe and timely control transfer is critical to explore human-machine collaboration in maritime operations. This study uses experimental data collected from autonomous ship Yujiaotou 001 (Figure 1), which is powered by a box-type battery-based green propulsion system and integrates ANS and ROC, and where manual operation by the onboard seafarer is required when the ship is unable to navigate independently under certain circumstances (such as in strong currents). To this end, it aims to investigate control transfer intentions of MASS onboard seafarers using behavioral data, proposing an intention prediction model using unsupervised machine learning and deep learning. By collecting body positions and hand features of seafarers and vessel trajectories, it first uses K-Means cluster analysis to identify indicators of different seafarer states and classify behaviors. It then combines supervised deep learning algorithms to achieve control transfer prediction of onboard seafarers. It provides a theoretical basis for constructing a more intelligent and safer MASS human-machine collaboration system.

Fig. 1. The autonomous ship Yujiaotou 001

The structure of the paper is organized as follows. Section 2 provides a literature review on human factors, risk assessment, and control transfer issues in MASS. The proposed methodology, including the clustering model and the CNN-LSTM-Attention prediction model, is demonstrated in Section 3. The experimental setup, data collection, and preprocessing procedures are detailed in section 4. Section 5 presents the clustering results, prediction performance, and corresponding discussion.

2. Literature review

In the field of control transfer for autonomous ships, existing research has focused on dimensions such as risk quantification, human factors influence, and transfer preparation time.

In the field of human factors engineering, Cheng et al. (2024) investigated human errors occurring during human-machine collaboration in MASS, revealing how factors such as available time, task complexity, and early warning mechanisms influence seafarers' takeover performance. Veitch et al. (2024) conducted simulation experiments featuring takeover and handover scenarios, and confirmed that factors such as multi-ship monitoring, time urgency, and the lack of Decision Support Systems (DSS) would affect the effectiveness of control transfer. This provides a basis for the design of the ROC and the training of seafarers.

To quantify the risks during the transfer process, frameworks such as Systems Theoretic Process Analysis (STPA), Hidden Markov Model (HMM) (Li et al., 2024), and Bayesian Belief Network (BBN) (Johansen and Utne, 2024) have been applied to risk modeling.

Furthermore, the research has recognized that the preparation time required for control transfer is not fixed but varies dynamically according to the complexity of the scenario. For inland vessels, a real-time decision support system based on the Situation-Operator Model (SOM) and Human Performance Reliability Score (HPRS) (Söffker et al., 2025) has verified that there are differences in the required cognitive preparation time in different situations.

Although these studies have provided insights into risk assessment and influencing factor analysis in MASS, a research gap in seafarers' control transfer intention remains unexplored. This study focuses on the control transfer intention by investigating human behavioral features, providing insights on MASS navigational safety.

3. Methodology

3.1. *Clustering*

The application of unsupervised machine learning models for data clustering analysis involves four steps: (1) Data pre-processing: the integrated dataset was normalized using Z-score standardization; (2) Optimal cluster determination: the Elbow Method, Silhouette Coefficient (SC), and Davies-Bouldin (DB) index were used to determine the optimal number of clusters (K) through cross-validation; (3) Feature refinement: highly correlated variables were excluded to reduce redundancy, and Principal Component Analysis (PCA) was applied for dimensionality reduction and noise suppression; (4) Clustering and analysis: the K-Means algorithm was applied to the refined dataset to partition the data. By analyzing cluster characteristics, differences, and centers, typical pre-transfer behavioral states of seafarers.

3.2. *Prediction model*

Long Short-Term Memory (LSTM) introduces a gating mechanism in Recurrent Neural Networks (RNN), regulating the flow and retention of information through input gates (i_t), forget gates (f_t), and output gates (o_t) (Lee and Park, 2024) . It effectively captures long-term dependencies by transmitting sequence information in the time dimension through cell states. In addition, LSTM often enhances the model's expressive power through multi-layer stacking, where each layer independently maintains its hidden state and cell state. The calculation formula is as follows (see Eq.(1)-(6)):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * \tanh(c_t) \tag{6}$$

σ denotes the sigmoid activation function and tanh represents the hyperbolic tangent function. $\tilde{c}_t$ signifies the candidate cell state, while c_t and h_t denote the cell state and hidden state at time t. x_t represents the input sequence value at the current time step, and W and b are the associated weight matrices and bias vectors.

The proposed CNN-LSTM-Attention model (Figure 2) integrates a Convolutional Neural Network (CNN), LSTM, and a self-attention mechanism. The framework is organized into six functional modules: the input layer, data preprocessing layer, feature extraction layer, attention mechanism layer, fully connected layer, and output layer.

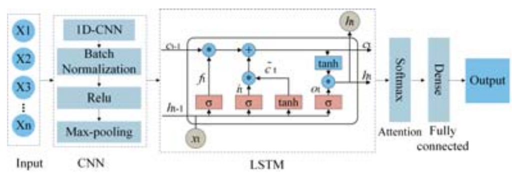

Fig. 2. The structure of CNN-LSTM-Attention

Initially, input data undergo normalization preprocessing. Subsequently, a CNN is utilized for preliminary feature extraction. Multiple convolutional kernels are applied to capture local features, while a sliding window mechanism is leveraged to extract local spatial dependencies within the sequence. The Relu activation function is then incorporated to enhance the nonlinear representational capacity of the extracted features. Furthermore, the LSTM layer is implemented to model and reinforce temporal dependencies. Following this, a self-attention layer is introduced to calculate the correlation between the features and trainable weight matrices. This mechanism enables the model to adaptively allocate attention weights based on data characteristics, thereby accentuating the influence of critical features on the final prediction. Finally, the extracted features are integrated and mapped through a fully connected layer to yield the definitive prediction results.

4. Experiment

4.1. *Experimental scenario*

To evaluate the effectiveness of the proposed method, this study utilized seafarer control transfer events recorded by the autonomous ship Yujiaotou 001 in the real waterway of the Shaying River in Henan Province as experimental data. The experimental scenarios are shown in Figure 3. Three control transfer events were recorded throughout the process, and the effective experimental duration was approximately 40 min.

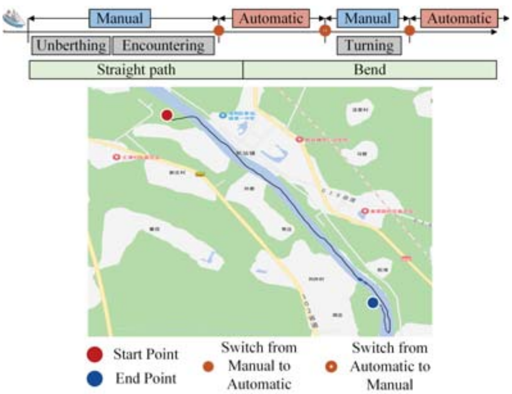

Fig. 3. Vessel control transfer events in different scenarios

As shown in Figure 3, the autonomous ship Yujiaotou 001 adopts manual control mode at the initial stage of leaving the berth and then meets the oncoming vessel. After the encounter, the seafarer adjusted the control mode from manual to automatic and guided the vessel into the S-bend. After passing the S-bend, the seafarer is ready to make a U-turn and switch the mode from automatic to manual. After completing the U-turn, during the return journey, the seafarer once again switched the control mode from manual to automatic.

4.2. *Experiment data*

4.2.1. *Data collection*

This study collected two categories of data: video data and vessel data. The data collection procedure is illustrated in Figure 4.

The video data is collected through camera installed in the wheelhouse. The collection process consists of three steps: first, the moment of control transfer was determined based on the seafarers' button-pressing action, serving as a benchmark for subsequent analysis. Second, a 30-second video segment preceding this moment was extracted as an analytical sample to focus on the seafarers' behavioral performance before the transfer. Finally, the video clips were processed using MediaPipe (Figure 5), an open-source deep learning framework that employs a two-stage detection-tracking strategy, where a lightweight CNN detector first locates the target and a tracker then follows keypoints across frames. This process enabled human pose estimation, from which the keypoint information of the seafarers' behavior was extracted.

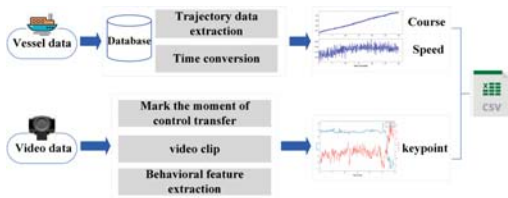

Fig. 4. Data collection procedure

Environmental calibration was performed: a reference frame with a clearly visible toggle button was selected through an interactive interface. The center coordinates of the button were manually labeled on this frame to establish the spatial correspondence between the image coordinates and the bridge. Based on the raw keypoints, two core behavioral features were engineered. First, the lean angle of the body (body_lean) was defined as the angle between the vertical axis and the line segment connecting the keypoints of the shoulder and hip. Second, a hierarchical distance estimation algorithm was developed to quantify the proximity of hand to the toggle button (dist_btn). Guided by the visibility confidence of keypoints, the algorithm prioritized the wrist keypoint (provided its confidence exceeded the set threshold) to compute the Euclidean distance to the toggle button. If the wrist was occluded, the elbow keypoint was automatically utilized as a proxy for the estimation. This strategy enhances the robustness of feature extraction under conditions such as partial occlusion.

Ultimately, all extracted features were precisely synchronized with the corresponding absolute timestamps and video frame numbers, organized into a structured time series dataset, thereby providing readily accessible quantitative behavioral data for downstream modeling and analysis.

Fig. 5. Keypoints extracted by MediaPipe

The vessel data is collected by the system and stored in a database. The primary objective of the data extraction phase is to filter and parse data packets from the original data table, thereby extracting vessel course and speed information. After time conversion, the algorithm retrieves vessel data corresponding to a predefined time window from the parsed data stream. Finally, trajectory data with precisely synchronized timestamps are exported, forming the basis for subsequent multimodal correlation analysis.

4.2.2. *Data processing*

To ensure data quality, this study conducted processing on the original data to eliminate noise and correct outliers.

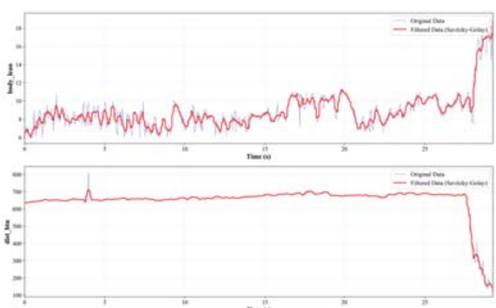

Fig. 6. Comparison of original data with filtered processing data

First, due to the precision limitations of the MediaPipe algorithm, the behavioral data obtained thereby may have the problem of keypoint jitter. Therefore, the Savitzky–Golay filter (Figure 6) (with a window width of 5 frames) was adopted to smooth the original data. This method achieves data smoothing through polynomial fitting within a sliding window, reducing noise while effectively preserving the main movement trends during the seafarers' actions.

Second, GPS devices may be affected by factors such as signal obstruction and equipment failure when collecting vessel trajectory data, which can lead to abnormal or missing values. Therefore, this study has verified the validity of the course data to ensure that all values are within the range of 0° to 360°. The inspection results show that no obvious abnormalities or missing values were found in the vessel trajectory data obtained in this study, so no further processing was carried out.

Finally, to establish the correlation between vessel trajectory data and seafarers' behavioral features, this study employs a linear interpolation method to perform time alignment and fusion of multi-source heterogeneous time series data. First, the time reference is unified: by parsing the timestamps from both data sources, the relative time is calculated, the temporal offset is determined, and the effective overlapping time interval is identified. Within this interval, independent one-dimensional linear interpolation functions are constructed for vessel speed and course, based on the high-frequency sampling points of the vessel trajectory data.

There are three reasons for choosing this method: (1) vessel trajectory data typically exhibit continuity and gradual variation over short time windows; (2) linear interpolation offers high computational efficiency, making it well-suited for large-scale time series processing; (3) at relatively high sampling frequencies, linear interpolation maintains sufficient accuracy for practical use.

The final output dataset comprises synchronized timestamps, interpolated vessel speed and course, as well as behavioral features extracted from the corresponding time frames. This constitutes a unified multimodal time series dataset, which serves as the foundation for subsequent correlation analysis.

5. Results and discussion

5.1. *Clustering results*

5.1.1. *Number of clusters*

Cluster analysis was performed based on the fused multimodal dataset of vessel course, speed, and video data. The analysis reveals (Figure 7) that at K=2, the Silhouette Score reaches its peak while the DB index attains its minimum value. This indicates that a binary clustering structure achieves the optimal equilibrium between intra-cluster compactness and inter-cluster separation.

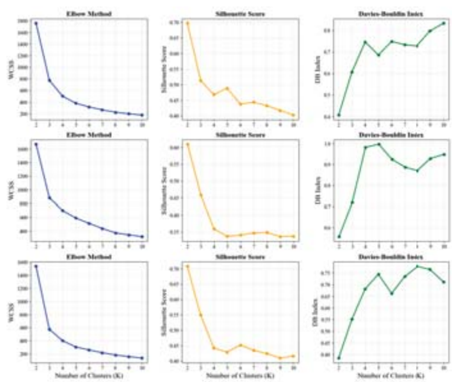

Fig. 7. Evaluation metrics for different numbers of clusters

Although K=2 does not coincide with the elbow point of the WCSS curve, the subsequent significant decline in the Silhouette Score and the marked increase in the DB Index suggest that further increasing the cluster count yields no substantial improvement in clustering quality. Instead, a higher K value may lead to overfitting and the fragmentation of meaningful behavioral patterns.

To further examine the actual discriminatory effectiveness of the clustering structure, PCA was applied for dimensionality reduction. The visualization of the clustering results is shown in the Figure 8. The two clusters have clear contours and good distribution in the two-dimensional space. The samples within the clusters are closely clustered, and there is no overlap between the clusters.

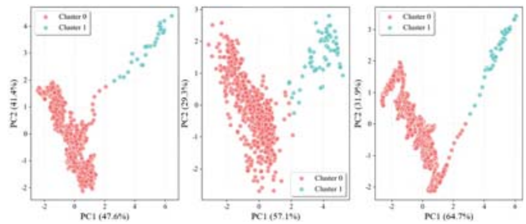

Fig. 8. The visualization of PCA dimensionality reduction

5.1.2. *Analysis of clustering results*

The clustering results based on the K-Means algorithm reveal two different behavioral patterns in the control transfer process of autonomous ships. The sample distribution exhibits imbalanced characteristics, where cluster 0 represents the predominant state, whereas cluster 1 characterizes a specific behavioral regime. Figure 9 shows the feature distribution of the boxplots of the two modes. A detailed analysis of these two behavioral patterns, integrated with the distribution characteristics observed in the boxplots, is provided below:

(1) Body_lean: in cluster 1 (control transfer preparation), both the mean and median values of the seafarers' body_lean are significantly elevated compared to cluster 0 (cruise), with a more concentrated data distribution. This intuitively reflects that as seafarers perceive potential risks and prepare to intervene, they will adopt a leaning-forward posture to obtain a better field of vision and facilitate faster reaction times.

(2) Dist_btn: the distance between the seafarers' hand and the toggle button is substantially reduced in cluster 1, with the distribution shifting downward and becoming more compact. This suggests that the proximity of the hand to the control device serves as the most direct and explicit precursor to executing a manual intervention.

(3) Vessel data: in the three described scenarios of control transfer, the speed and course of the vessels in cluster 1 showed very little change. On the contrary, the vessels in cluster 0 observed more significant changes in course, which is consistent with the logic of making course corrections either at the end of the encounter or before initiating the turn. Moreover, the numerical difference between

these two categories is very small.

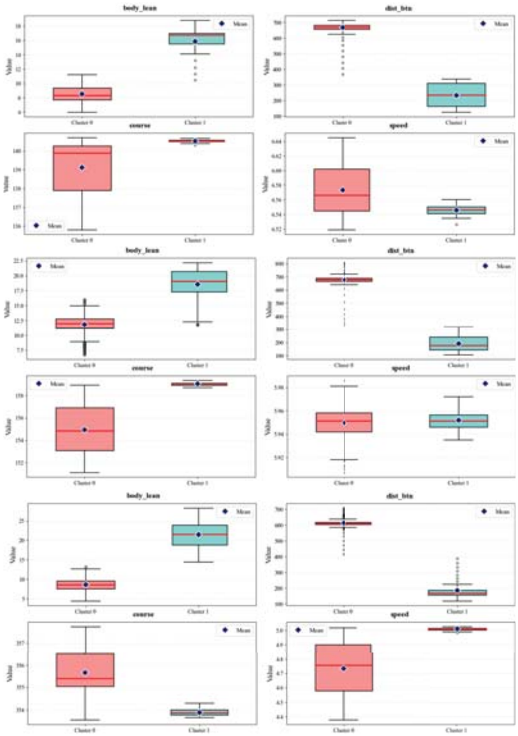

Fig. 9. Feature distribution boxplots of two modes

(4) Synthesis: a comprehensive comparison of the two clusters indicates that the marked discrepancies in dist_btn and body_lean, specifically the significant reduction in distance and the increase in forward leaning, establish a boundary for distinguishing "cruise" from "control transfer preparation". The high degree of consistency in vessel speed and course across both clusters suggests that, within the current observation window, control transfer shifts do not manifest immediately in trajectory metrics; instead, they are primarily reflected in the body adjustments of the seafarer.

5.2. *Prediction results and analysis*

The number of hidden units in the LSTM layer was set to 64. The hyperparameters for model training were configured as follows: a learning rate of 0.001, a batch_size of 32, and a total of 50 training epochs. The Adam optimizer was employed to facilitate weight updates and convergence. The integrated dataset was partitioned into training, validation, and testing sets according to a ratio of 7:2:1, ensuring a rigorous evaluation of the model's generalization capabilities. The proposed model achieved an overall accuracy of 96.40% and an F1-score of 96.29% on the test set. The comparison between the ground truth and the predicted values is illustrated in Figure 10.

Fig. 10. The comparison between the true value and the predicted value

6. Conclusion

When autonomous ship is unable to navigate autonomously, onboard seafarers need to assume manual operation to reduce risks. This study demonstrates that control transfer intention is not merely an unobservable internal psychological state; rather, it can be characterized and predicted through manifest and quantifiable body postures.

Based on the K-Means clustering analysis, seafarers' behavior was categorized into two distinct regimes: "cruise" and "control transfer preparation". The two regimes exhibit significant statistical discrepancies across the identified metrics, providing a foundation for intention prediction based on behavioral states. To enhance the capabilities of the prediction model, a CNN-LSTM-Attention model was developed. By incorporating a self-attention mechanism, the model focuses on critical behavioral fragments and strengthens the learning of valuable information within complex temporal dependencies. The resulting high predictive accuracy validates the feasibility of control transfer intention recognition derived from behavioral data.

This research offers novel insights into the safety of human-machine collaboration within autonomous ships, facilitating the development of intelligent bridge systems equipped with proactive early-warning and risk prevention capabilities. Furthermore, the limitations of this study is the small sample size. In practice, delayed takeover, whether caused by communication latency or the seafarers' absence from the bridge, can impact the applicability of intention models. Future work will focus on enhancing model generalization by expanding the dataset to include behavioral characteristics under different monitoring modes, as well as additional data types such as physiological and environmental factors. Such advances will drive the development of a more safe and reliable human-machine collaboration framework for autonomous ships.

References

Cheng, T., E. A. Veitch, I. B. Utne, M. A. Ramos, A. Mosleh, O. A. Alsos, and B. Wu (2024). Analysis of human errors in human-autonomy collaboration in autonomous ships operations through shore control experimental data. *Reliability Engineering & System Safety 246*, 110080.

Fan, C., K. Wróbel, J. Montewka, M. Gil, C. Wan, and D. Zhang (2020). A framework to identify factors influencing navigational risk for maritime autonomous surface ships. *Ocean Engineering 202*, 107188.

Feng, M., C. Chen, L. Jialun, W. Xuming, and Y. Xinping (2022). Key technologies of ship remote control system in inland waterways under ship-shore cooperation conditions. *Chinese Journal of Ship Research 17*(5), 125–133.

Johansen, T. and I. B. Utne (2024). Human-autonomy collaboration in supervisory risk control of autonomous ships. *Journal of Marine Engineering & Technology 23*(2), 135–153.

Lee, H.-J. and D.-J. Park (2024). Collision evasive action timing for mass using cnn–lstm-based ship trajectory prediction in restricted area. *Ocean Engineering 294*, 116766.

Li, W., W. Chen, Y. Guo, S. Hu, Y. Xi, and J. Wu (2024). Risk performance analysis on navigation of mass via a hybrid framework of stpa and hmm: Evidence from the human–machine co-driving mode. *Journal of Marine Science and Engineering 12*(7), 1129.

Li, X. and K. F. Yuen (2024). A human-centred review on maritime autonomous surfaces ships: impacts, responses, and future directions. *Transport Reviews 44*(4), 791–810.

Munim, Z. H., T. Notteboom, H. Haralambides, and H. Schøyen (2025). Key determinants for the commercial feasibility of maritime autonomous surface ships (mass). *Marine policy 172*, 106482.

Söffker, D., A. Bejaoui, and O. Shyshova (2025). Progress towards multidimensionally scalable assisted and/or automated ship navigation and control–part ii: human in the interaction loop. *Journal of Marine Engineering & Technology*, 1–11.

Veitch, E., O. A. Alsos, T. Cheng, K. Senderud, and I. B. Utne (2024). Human factor influences on supervisory control of remotely operated and autonomous vessels. *Ocean Engineering 299*, 117257.

Wróbel, K., K. Formela, P. Kołakowski, and M. Gil (2025). Time budget for merchant ship control takeover-preliminary results. *TransNav, International Journal on Marine Navigation and Safety od Sea Transportation 19*(3), 781–786.

Yan, X.-p., C. Li, J.-l. Liu, X. You, S.-w. Wang, and F. Ma (2021). Architecture and key technologies for new generation of waterborne transportation system. *Journal of Transportation Systems Engineering and Information Technology 21*(5), 22.