

Improving algorithms for artificial intelligence-based crowd density estimation and public safety assurance

F.Sh.Atamuratov

Inha University in Tashkent

O.O.Rajabov

University of Public Safety of Republic of Uzbekistan

Abstract: *The article examines the automatic estimation of crowd density from video imagery in places of mass gathering, as well as the task of converting the resulting estimate into a management decision aimed at ensuring public safety. The stages in the development of density estimation algorithms are analysed comparatively, and the limitations of modern deep-learning-based solutions in terms of accuracy, computational resources and adaptability to other environments are demonstrated on the basis of sources published in 2025-2026. A four-stage algorithmic scheme is proposed, comprising modules for camera calibration, density map generation, risk level classification and the issuing of recommendations on force deployment, and its operation is demonstrated by means of a numerical example. The conditions for implementing the system within the current legal and regulatory requirements of the Republic of Uzbekistan in the field of artificial intelligence are also substantiated.*

Keywords: *crowd density, density map, convolutional neural network, risk level, intelligent video analytics, public safety, decision support, camera calibration, domain shift*

INTRODUCTION

The cause of the tragedies that occur at mass events, places of pilgrimage, transport hubs and shopping complexes is often not, as is commonly assumed, the “panic” of the crowd, but rather the limited nature of the available space. Once the number of people exceeds a certain threshold, a crowd begins to move not as a sum of separate individuals but as a continuous medium: an impulse at one point turns into a pressure wave several metres away, and a person who has fallen loses the ability to get back up. The principal mechanism of injury and death is compressive asphyxia resulting from compression of the chest. An important aspect of this process is that it is directly related to a measurable physical parameter - the number of people per unit area. It is therefore technically possible to move from a subjective assessment of risk to quantitative control.

In practice, this task has so far been assigned mainly to human attention. A duty operator monitors dozens of screens, but because of the natural limits of attentional resources the gradual growth of density - precisely the most dangerous, slowly developing scenario - is often noticed too late. As video surveillance networks expand, this imbalance increases: the number of cameras grows, while the number of staff analyzing them remains almost unchanged. For this reason, the automatic analysis of video streams and the early signaling of a dangerous situation are becoming increasingly relevant.

In Uzbekistan this direction has been consolidated at the level of state policy. The Strategy approved by Presidential Resolution No. PR-358 of 14 October 2024 establishes as target indicators the increase of the volume of software products and services created on the basis of artificial intelligence to 1.5 billion US dollars, an increase of 10 in the number of scientific laboratories operating in this field, and securing a place among the top 50 countries in the Government Artificial Intelligence Readiness Index [1]. Presidential Decree No. PD-189 of 22 October 2025 linked these

tasks to specific deadlines: the implementation of at least 100 projects in public administration, the social sphere and sectors of the economy by the end of 2026, the attraction of more than 1 billion US dollars of foreign investment into the field by the end of 2030, and the launch of 15 artificial intelligence laboratories in higher education institutions by the end of 2026 [2]. At the same time, the list of priority projects approved by Resolution No. 425 of the Cabinet of Ministers of 10 July 2025 provides for the introduction of a video identification system using computer vision in order to ensure public safety, monitor places of mass gathering and detect crimes (deadline for developing the information system - October 2026), as well as the creation of a system for monitoring the number and activity of visitors at entertainment and retail facilities with the help of artificial intelligence (November 2026) [3]. An institutional base has also been formed: it has been reported that a unified “Safe City” platform integrated with nearly forty departmental databases has been created in the practice of the internal affairs bodies, and that more than 20 new information systems have been launched [6].

Thus, both the technical infrastructure and the legal basis are in place. However, an analysis of the scientific literature reveals one systemic gap: most existing solutions are confined to estimating the size or the density of a crowd and do not convert the result into a management recommendation that is intelligible to a commander and organised step by step - that is, into a justified proposal indicating where, how many and when forces should be deployed. In addition, high accuracy figures are, as a rule, maintained only within the dataset on which the model was trained and decline appreciably in a different environment.

The aim of the article is to analyse critically the current state of algorithms for estimating crowd density by means of artificial intelligence and to substantiate an algorithmic scheme that covers the entire chain from density estimation to recommendations on the allocation of forces and that complies with legal and ethical requirements. The study employs the methods of systems analysis, comparative analysis, mathematical modelling and the interpretation of legal and regulatory documents; the empirical basis consists of publications indexed in the Scopus and Web of Science databases in 2025–2026 and the current regulatory documents of the Republic of Uzbekistan.

MAIN PART

1. The density indicator and its risk thresholds

The principal measurable quantity is the local density:

$$\rho = N / S_{\text{eff}}, [\text{persons}/\text{m}^2] \quad (1)$$

where N is the number of people in the selected zone, S_{eff} is the effective area of that zone, that is, the value obtained by subtracting columns, flower beds, retail kiosks, barriers and other occupied surfaces from the total area. The most widespread error encountered in practical calculations is the use of the total area. For example, if 300 m² of an area of 1,500 m² is occupied by structures, then for 5,400 people the calculation based on the total area yields 3.6 persons/m², whereas the calculation based on the effective area yields 4.5 persons/m². The difference amounts to 20 per cent, and it falls precisely within the interval that separates the safe and the dangerous regimes.

The threshold values established in classical studies of crowd dynamics (the classification of levels of service proposed by J. Fruin and the work of G. K. Still) are summarised in practice in the following scale (Table 1). This scale has been supplemented by the author with management measures.

Table 1.

Crowd density levels and the corresponding management measures

Level	ρ , persons/m ²	State of the crowd	Management measure
I	< 1.0	Free movement, unobstructed passage	Routine monitoring
II	1.0–2.0	Movement speed decreases slightly	Enhanced monitoring, control of the direction of flow
III	2.0–4.0	Movement slows down, physical contact begins	Regulation of the inflow, redirection of reserves
IV	4.0–5.0	Independent movement becomes difficult, the risk of falling increases	Restriction of entry, deployment of additional forces, warning
V	> 5.0 (above 6.0 - critical)	The crowd moves as a single medium, risk of compressive asphyxia	Suspension of entry, staged evacuation

The main advantage of the scale is its operational character: each level corresponds to a specific action, so that the number produced at the output of the algorithm is connected directly to the service protocol.

2. The logic behind the development of density estimation algorithms

First-generation approaches were based on the detection of individual people. In sparse crowds they produce good results and return the coordinates of each person, but in dense scenes accuracy falls sharply because bodies occlude one another. The second generation - regression methods - estimates the total count directly from image features; such methods are resistant to occlusion but provide no information about the spatial distribution of people, so it remains unknown at which point a build-up has occurred.

The concept of the density map emerged as the third generation. Here each head point is spread by means of a Gaussian kernel:

$$D(x) = \sum \delta(x - x_i) * G_{\sigma}(x) \quad (2)$$

The model reconstructs the density map from the image, and the total count is obtained as its sum:

$$\hat{N} = \sum_x \sum_y \hat{D}(x, y) \quad (3)$$

The mean absolute error and the mean squared error are used as quality criteria:

$$MAE = (1/M) \sum |\hat{N}_i - N_i|, \quad MSE = (1/M) \sum (\hat{N}_i - N_i)^2 \quad (4)$$

It was within this paradigm that multi-column (MCNN) and single-column dilated-convolution (CSRNet) architectures appeared, followed by point-supervised and transformer-based models. All of them contend with the same four problems: scale variation (the sizes of heads close to and far from the camera differ sharply), occlusion, perspective distortion and variability of illumination.

From the point of view of practical deployment, a separate direction is the economy of computational resources. In a study published by Qatari researchers in 2025, three lightweight models intended to run on edge devices, with no more than 0.2 million parameters, were trained by means of knowledge distillation; LCDnet has only 0.05 million parameters, MCNN 0.13 million and DroneNet 0.15 million, and CSRNet served as the teacher model for them [7]. That work reports that the inference time of the MCNN and DroneNet models was reduced 8-fold relative to the teacher model and 18-fold in the case of LCDnet, and that the required memory decreased from 65 MB to less than 1 MB; the gain obtained from distillation was highest when 40 per cent of the labelled data was used and lowest when 100 per cent was used [7]. This last observation is of

particular value for local conditions: distillation yields the greatest effect at a stage when the national dataset is still small.

The fourth direction is the transition from static density to motion dynamics. In a study published in 2026, the state of a crowd was divided into four levels - normal, moderate, dense and dangerous - on the basis of dense optical flow and a hybrid CNN–LSTM architecture; the authors consider such graded classification preferable to a binary “anomaly present/absent” signal, because it makes it possible to organise intervention in stages - warning, redirection of the flow, partial evacuation [8]. This approach takes into account, together with density, its rate of change, the divergence of the flow and the motion entropy.

3. Structure of the improved algorithm

On the basis of the analysis above, a four-module scheme is proposed. Its composition is shown in Figure 1.

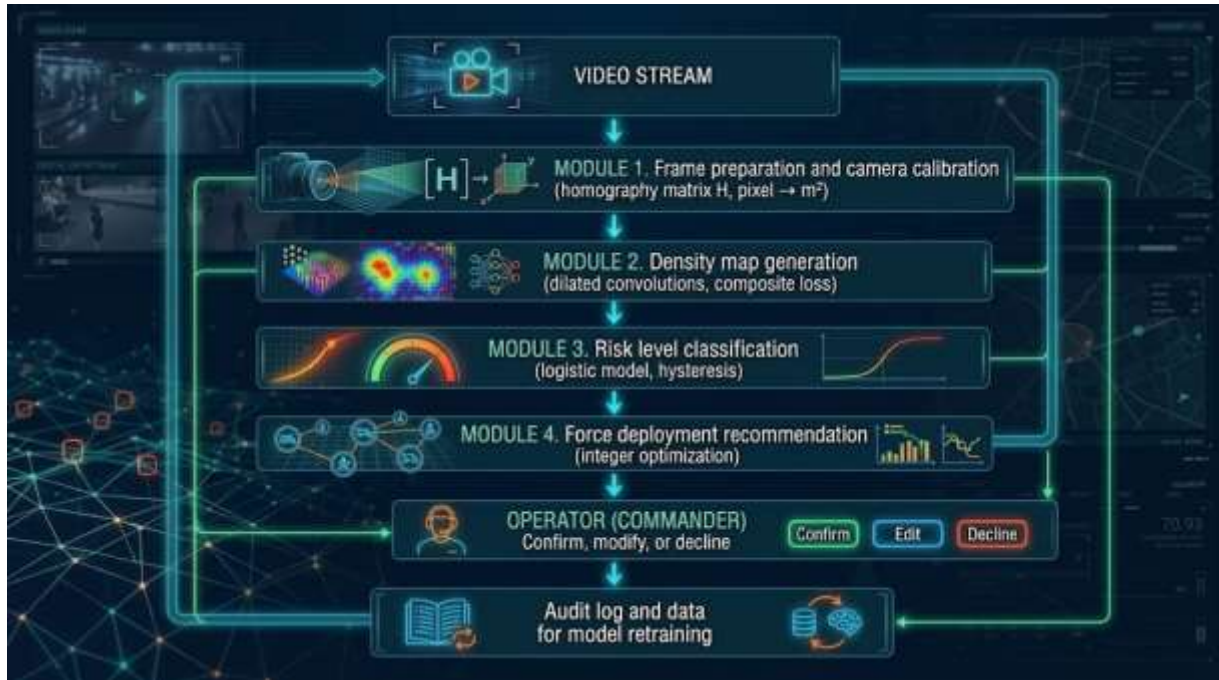


Figure 1. Block diagram of the proposed system

Module 1 - calibration. The density map itself is expressed in pixels, whereas safety requires a value per square metre. For this reason, a homography matrix that maps the image plane onto the ground plane is determined for each camera, and each pixel is assigned the weight $w(x, y)$ of its real area. As a result, the density of a zone is computed as follows:

$$\hat{\rho}_k = \sum \hat{D}(x, y) / \sum w(x, y), (x, y) \in \Omega_k \quad (5)$$

Without this stage the system returns a number, but a number without a unit of measurement - a situation that is unacceptable in matters of safety.

Module 2 - density map. A VGG-16-based feature extractor and a layer of dilated convolutions are used: such a scheme widens the receptive field without sharply reducing the spatial dimensions. A composite loss function is applied during training:

$$L = L_{\text{pixel}} + \lambda_1 L_{\text{count}} + \lambda_2 (1 - \text{SSIM}(\hat{D}, D)) \quad (6)$$

The first term ensures the pixel-by-pixel accuracy of the map, the second the correctness of the total count, and the third the preservation of the structure of the map. The significance of the third term is that a map with a correct total count but an incorrectly indicated location of the build-up is useless for management purposes.

Module 3 - risk level classification. The input feature vector: the density ρ , its rate of growth $\Delta\rho/\Delta t$, the mean speed of the flow $\bar{v}$ and the motion entropy E . The probability of risk is estimated by means of a logistic function:

$$P(\text{risk}) = 1 / (1 + e^{-z}) \quad (7)$$

where $z = \beta_0 + \beta_1\rho + \beta_2(\Delta\rho/\Delta t) + \beta_3\bar{v} + \beta_4E$.

In order to prevent the signal from “flickering” near the thresholds, hysteresis is introduced: the level changes only when the new value has been confirmed in three consecutive time windows. This reduces the number of false alarms and preserves the confidence of the operator.

Module 4 - recommendation. The allocation of forces is posed as an integer optimisation problem:

$$\min \sum c_j x_j; \sum a_{ij} x_j \geq r_i; \sum x_j \leq X_{\max}; x_j \in \mathbb{Z}_{\geq 0} \quad (8)$$

where x_j is the number of personnel drawn from group j , c_j is the cost of relocating them (distance, time), r_i is the requirement of zone i arising from its risk level, and $X_{\max}$ is the available reserve. The output is not an order, but a justified proposal.

A fundamental constraint. The system is not a decision-making entity. This requirement is not merely advisory in nature but is enshrined in regulation: according to the current rules of ethics, participants in artificial intelligence may not accept decisions taken by a system as final decisions; all decisions must be under human supervision and subject to human review, and the final decision must be taken by a human being [4].

4. Worked numerical example

Let us assume that the effective area of a square is $S_{\text{eff}} = 1,200 \text{ m}^2$ and that the model has estimated 5,400 people. Then, according to formula (1), $\rho = 4.5 \text{ persons/m}^2$ - level IV of Table 1. At the threshold of 4.0 persons/m^2 the capacity of the zone amounts to 4,800 people, which means that the current build-up exceeds the norm by 600 people, or by 12.5 per cent.

Let us take the dynamics into account: if over the last 5 minutes the number of people has grown from 4,600 to 5,400, the net intensity of the inflow is 160 people per minute. The time remaining until the critical level of 6.0 persons/m^2 (7,200 people) is reached is:

$$(7,200 - 5,400) / 160 \approx 11 \text{ minutes.}$$

This is a real time window for management. It is precisely within these 11 minutes that entry must be restricted, an additional passageway opened and reserves relocated. If the rate of entry is reduced to 160 people per minute, that is, if entry and exit are balanced, the growth of density will cease.

The requirement for forces is computed by means of model (8). If the conditional coefficient set by the commander for level IV is taken as 1 officer per 100 people, the requirement amounts to 54 officers; if 30 officers are present on site, the shortfall is 24 units, and the system proposes the option of drawing on the two nearest reserve groups.

Finally, let us assess the influence of the model error. With $\text{MAE} \approx 60$ people, the relative error in a scene of 5,400 people amounts to 1.1 per cent, and in terms of density to $60/1,200 = 0.05 \text{ persons/m}^2$ - twenty times smaller than the width of a level (1.0 persons/m^2), which means that the classification is stable. If the model is poorly adapted to another environment and the MAE rises to 300 people, the error reaches 0.25 persons/m^2 ; this is still smaller than the width of a level, but at values close to a threshold (around 4.0 or 5.0) the probability of misclassification increases appreciably. It is precisely for this reason that hysteresis and the local calibration of thresholds are a mandatory rather than an optional element of the algorithm.

ANALYSIS AND DISCUSSION

5.1. The balance between accuracy and feasibility of implementation

While accuracy is the priority criterion in scientific publications, in practice the decisive factor is the ability of a model to run on the existing infrastructure. Table 2 presents the resource indicators of models designed for edge devices.

Table 2.

Resource requirements of lightweight models and the calculations derived from them

Indicator	Value	Author's calculation
LCDnet parameters	0.05 M	3 times fewer than DroneNet
MCNN parameters	0.13 M	-
DroneNet parameters	0.15 M	-
Reduction of inference time (MCNN, DroneNet)	8-fold	a decrease of 87.5 %
Reduction of inference time (LCDnet)	18-fold	a decrease of 94.4 %
Memory footprint: teacher → student model	65 MB → < 1 MB	a saving of more than 98.5 %

These figures lead to a practical conclusion: instead of increasing the capacity of a centralised server without limit within an existing camera network, it is expedient to perform the initial analysis on an edge device. In that case it is not video that is transmitted over the network, but compact metadata on the density and the risk level - which reduces both the load on the channel and the risk associated with personal data.

5.2. Domain shift as the principal practical risk

The high accuracy figures published in the literature should be approached with caution. Table 3 demonstrates this clearly.

Table 3.

Results of risk level classification (CNN–LSTM and optical flow model)

Indicator	Value
Number of classes	4 (normal, moderate, dense, dangerous)
Training and validation set	10,000 annotated frames
Validation accuracy	99.23 % (258 of 260 samples correct)
Nature of the errors	Both errors lie on the “moderate–dangerous” boundary
AUC	1.000 for three classes; 0.9994 for the “dangerous” class
Test on another dataset (UMN)	7 of 10 sequences correct (70 %)
Completeness of anomaly detection (recall)	80 %
Latency	0.173 s for a 16-frame window (≈ 12.77 fps, CPU only)

Although the model produced almost perfect results within its own domain, on the UMN dataset, which did not take part in training, it correctly identified four of five anomalous sequences and three of five normal sequences; the authors explain this by differences in camera angle, scene structure and illumination [8]. According to the author's calculation, the overall accuracy fell from 99.23 per cent to 70 per cent, that is, by 29.2 percentage points. This figure yields the central conclusion for local implementation: a model trained on a foreign dataset cannot be transferred directly to the conditions of Uzbekistan. The necessary measures are as follows: the formation of a local dataset from the “Safe City” cameras, separate calibration and re-tuning of the threshold values for each camera, and the introduction of a procedure for the periodic revalidation of the model.

The latency figure yields another important conclusion: although a response time of 0.173 seconds falls short of frame-by-frame real-time operation, it is entirely sufficient for early warning, because a dangerous growth of density develops over minutes rather than seconds. The 11-minute margin shown in the calculation example above is clear evidence of this.

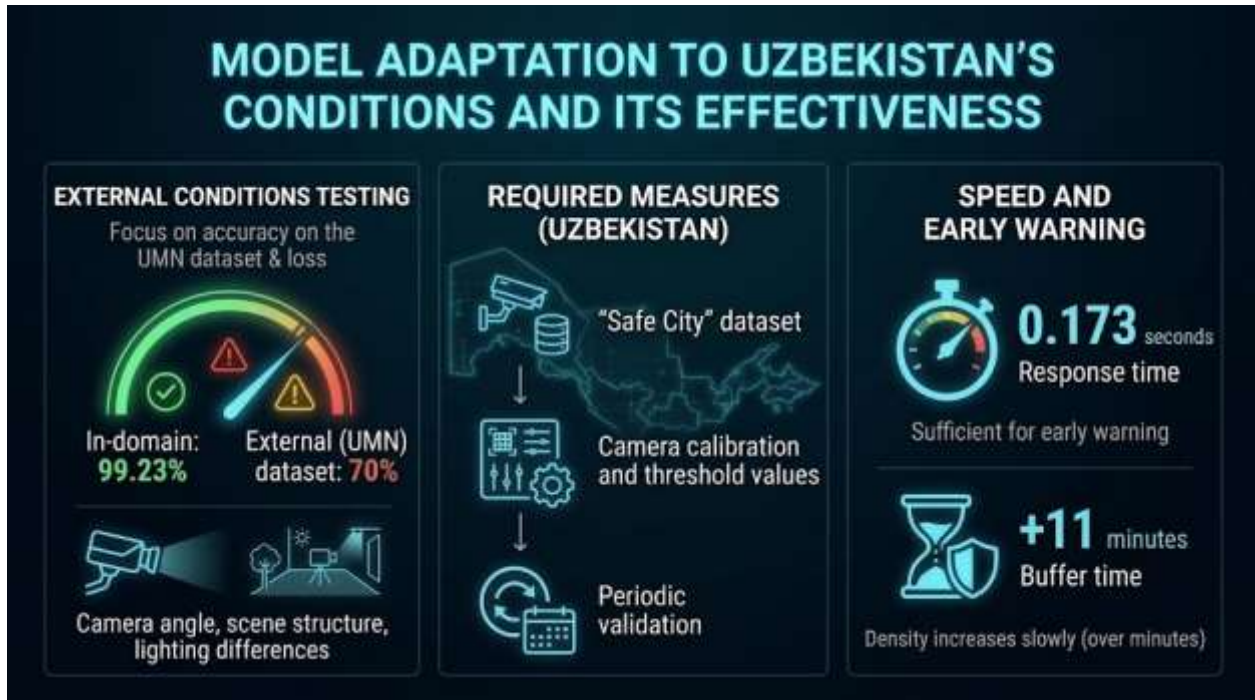


Figure 2. Model localisation and performance

5.3. The question of scale: not every camera, but the right point

At large-scale events the artificial intelligence layer is deployed selectively. At the Maha Kumbh Mela in 2025, where more than 400 million participants were expected, more than 2,760 video surveillance cameras were used, of which 500 had the capability of real-time density analysis and predictive analytics [8]. According to the author's calculation, this constitutes approximately 18 per cent of all cameras. Consequently, effectiveness is ensured not by making the entire set of cameras "intelligent", but by correctly selecting entry and exit points, narrow passageways and places where flows intersect. Priorities under local conditions should likewise be determined according to this principle.

5.4. Legal and ethical boundaries

The two technological ways of monitoring places of mass gathering are not equivalent: in legal terms they fall under different regimes (Table 4).

Table 4.

Comparative legal and ethical profile of the two approaches

Criterion	Density map (anonymous counting)	Identification through face recognition
Data processed	Aggregated distribution; the individual is not identified	Biometric personal data
Personal data regime	Not applicable or minimal	Fully applicable [5]
Principal task	Prevention of risk, control of capacity	Identification of a wanted person, recording of an offence
Consequence of an error	False alarm, excessive deployment of forces	Misidentification of a person, direct impact on rights
Feasibility of running on	High (model size < 1 MB) [7]	Limited

Criterion	Density map (anonymous counting)	Identification through face recognition
an edge device		

It is precisely for this reason that a module based on a density map is preferable: it estimates the overall distribution rather than a specific individual and creates less risk with respect to the requirements of the Law “On Personal Data” [5]. At the same time, the rules of ethics impose additional requirements on the system: according to the principle of intelligibility, the operating logic of the system, its principal algorithmic criteria and its decision-making process must be presented in an explainable form, while users must be informed that the results may be used only in an advisory or informational capacity; the principle of fairness requires that no discrimination be permitted on the grounds of sex, race, ethnicity, language or social status [4]. In practice this means that the system must display, together with each alert, the grounds for it - the density value, the rate of growth and the accuracy interval - and that all recommendations and the decisions taken on their basis must be recorded in an audit log.

5.5. Linking scientific results to practice

The deadlines established in accordance with Resolution No. 425 of the Cabinet of Ministers (October and November 2026) [3] create a clear time window for scientific developments in this field: methodological solutions, a local dataset and evaluation metrics will acquire practical value only if they are incorporated into the technical specifications of precisely these projects. The question of personnel and infrastructure can be resolved through the laboratories being established under Decree No. PD-189 [2].

The limitations of the study should also be stated openly. The proposed scheme has been developed for daylight illumination and a fixed camera geometry; under conditions of night-time, heavy precipitation and extremely dense build-up, additional modalities (thermal imaging, multi-camera fusion) are required. In addition, assessing model quality by MAE alone is not sufficient: in matters of safety the principal criteria must be the proportion of correct classifications by level, the frequency of false alarms and the number of missed dangerous situations.

Conclusion

In ensuring public safety, crowd density should be regarded not as a subjective assessment but as a continuously measurable physical parameter corresponding to the number of people per unit area. Real-time monitoring of density makes it possible to move from a reactive approach that follows the emergence of a danger to a preventive approach based on early detection. Modern algorithms are evolving from the simple detection of people to the generation of density maps and the analysis of motion dynamics. Here, alongside the accuracy of a model, the possibility of running it efficiently on edge devices is also important: knowledge distillation makes it possible to reduce the size of a model by more than 98 per cent and to accelerate the inference process 8- to 18-fold. At the same time, the sharp decline of accuracy in different environments demonstrates the problem of domain shift and makes it necessary to adapt and calibrate foreign models to local conditions. The density estimation module must operate within a single technological chain together with camera calibration, risk level classification and recommendations on the allocation of forces and resources. Monitoring the rate of growth of density creates the possibility of early warning more than 10 minutes in advance, and this time serves to regulate the inflow and to direct reserve forces to the relevant area. Such a system must be applied not as the entity that takes the final decision, but as a tool that provides advisory information to a human operator, the decision itself being taken by a human being. The use of anonymous density maps is more consistent with the requirements of

privacy and personal data protection than approaches based on the identification of individuals. For this reason, under the conditions of Uzbekistan, the formation of a dedicated dataset on the basis of local video surveillance data, the adaptation of risk thresholds to the geometry of facilities and the step-by-step introduction of artificial-intelligence-based systems into practice constitute a pressing task.

References

1. Resolution of the President of the Republic of Uzbekistan No. PR-358 of 14 October 2024 “On approval of the Strategy for the development of artificial intelligence technologies until 2030” // National Database of Legislation. – <https://lex.uz/docs/-7158604>
2. Decree of the President of the Republic of Uzbekistan No. PD-189 of 22 October 2025 “On additional measures for the further development of artificial intelligence technologies” // National Database of Legislation, 27.10.2025, No. 06/25/189/0978.
3. Resolution of the Cabinet of Ministers of the Republic of Uzbekistan No. 425 of 10 July 2025 “On measures for the implementation of priority projects in the field of artificial intelligence technologies in 2025–2026” // National Database of Legislation, 11.07.2025, No. 09/25/425/0597.
4. Rules of ethics for the creation, implementation and use of solutions based on artificial intelligence: approved by Order No. 284-mh of the Minister of Digital Technologies of the Republic of Uzbekistan of 24 February 2026 (registered with the Ministry of Justice on 14 March 2026 under No. 3787) // National Database of Legislation, 16.03.2026, No. 10/26/3787/0237.
5. Law of the Republic of Uzbekistan No. \LRU-547 of 2 July 2019 “On Personal Data” // National Database of Legislation. – <https://lex.uz/docs/-4396419>
6. Congratulatory address of the President of the Republic of Uzbekistan to the officers and veterans of the internal affairs bodies, 25 October 2025. – <https://president.uz/oz/lists/view/8593>
7. Khan M.A., Menouar H., Hamila R., Abu-Dayya A. Crowd counting at the edge using weighted knowledge distillation // Scientific Reports. – 2025. – Vol. 15. – Art. 11932. – DOI: 10.1038/s41598-025-90750-5
8. Bhola G., Srivastava S., Nidhan H.R., Kathed S. Stampede detection and crowd analysis using CNN-LSTM and Farneback optical flow // Scientific Reports. – 2026. – Vol. 16. – Art. 16155. – DOI: 10.1038/s41598-026-45262-1
9. Jumaniyozov, T. Sh. OLIY HARBIY TA’LIM MUASSASALARIDA DARS JADVALINI AVTOMATLASHTIRISH TIZIMINI ISHLAB CHIQISH MUAMMOLARI // ORIENSS. 2023. №21. URL: <https://cyberleninka.ru/article/n/oliy-harbiy-ta-lim-muassasalarida-dars-jadvalini-avtomatlashtirish-tizimini-ishlab-chiqish-muammolari> (дата обращения: 07.10.2026).