UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Computer Science
Computer Science Curriculum

Mykyta Baliesnyi

Energy-Based Models for End-to-End Autonomous Driving

Master's Thesis (30 ECTS)

Supervisor(s): Tambet Matiisen, MSc

Tartu 2022

Energy-Based Models for End-to-End Autonomous Driving

Abstract:

Energy-based models (EBMs), a promising class of machine learning models, have shown impressive results in several domains, from natural language generation to computer vision. Learning to imitate expert demonstrations using an EBM has recently achieved state-of-the-art results in robotics, made possible by EBMs' better ability to handle multimodal probability distributions and learn behavior with abrupt command changes. In this work, EBMs are tested for the first time in the end-to-end autonomous driving domain on a real car. As a result, it is discovered that a simple EBM variant performs slightly better and is more stable than a baseline conventional neural network architecture. At the same time, EBMs turn out to exhibit a higher variability of predictions over time, or whiteness. As a solution to this problem, this work introduces a regularization technique that makes the predictions more smooth over time. In addition, an energy-based uncertainty metric is proposed, but its usefulness could not be assessed with sufficient reliability due to an insufficient number of real car evaluations. The thesis suggests several ideas for future work, such as using a different sampling method and comparing against mixture density networks.

Keywords:

end-to-end, autonomous driving, neural networks, behavioral cloning, energy-based models, real car

CERCS:

P170, Computer science, numerical analysis, systems, control

Energiapõhiste mudelite kasutamine närvivõrkudel põhineva isejuhtimise jaoks

Lühikokkuvõte:

Energiapõhised mudelid (EPM) on näidanud muljetavaldavaid tulemusi mitmes valdkonnas, näiteks loomuliku teksti genereerimine või masinnägemine. Käitumise kloonimine kasutades EPM-i roboti käitumisreeglite esitamiseks saavutas hiljuti muljetavaldavaid tulemusi, eelkõige tänu oma võimele käsitleda multimodaalseid tõenäosusjaotusi ja roboti käskluste järske üleminekuid ühelt väärtuselt teisele. Selles töös testitakse esimest korda EPM-e päris auto autonoomseks juhtimiseks. Tulemusena selgus, et EPM-i lihtne variant toimib pisut paremini ja on stabiilsem kui võrdlusena kasutatud tavapärane närvivõrgu arhitektuur. Samas selgus, et EPM-idel on suurem ennustuste varieeruvus ehk *whiteness*. Selle probleemi lahenduseks pakutakse töös välja regulariseerimisvõte, mis muudab erinevate ajasammude ennustused ühetaolisemaks. Lisaks pakutakse välja energiapõhine määramatuse mõõdik, kuid selle kasulikkust ei õnnestunud piisava usaldusväärsusega hinnata ebapiisava arvu sõitude tõttu päris autoga. Töö pakub välja mitmeid ideid tulevasteks edasiarendusteks, näiteks teistsuguse valimi valimise meetodi kasutamine ja võrdluse segujaotuse ennustamisega.

Võtmesõnad:

täielikult närvivõrkudel põhinev juhtimine, isejuhtivad autod, närvivõrgud, käitumise kloonimine, energiapõhised mudelid, päris auto

CERCS:

P170, Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimisteooria)

Acknowledgements

First, I would like to thank my fantastic supervisor, Tambet Matiisen, for his impeccable guidance. He sparked my interest in autonomous driving and inspired this project. Working alongside him was a great pleasure and the highlight of my academic career.

Second, a special thanks go to Romet Aidla for sharing the baseline model training pipeline he used for his thesis and answering my numerous questions about the prior work he authored. His openness to help made this thesis much easier to complete.

Third, I am thankful to Pete Florence and Igor Mordatch for discussing energy-based models.

Next, I am very grateful to Kertu Toompea, and especially Muhammad Zain Bashir, for spending their days as my safety drivers in the middle of an already so short Estonian summer.

I would like to thank my friends for their help keeping me on track while working on the thesis. You know who you are.

Finally, I would like to thank my beloved mom, dad, and grandmother, who have asked about how my thesis was doing many more times than I thought about it myself.

Contents

1 Introduction

Energy-based models (EBMs) assign a scalar value (energy) to possible states. For example, given a chess board and a set of candidate moves, an EBM could produce energy values for the states to which the candidate moves lead. Lower energy means "better" states. EBMs have been shown to have several theoretical and practical benefits over standard models in natural language, computer vision, and robotics domains [8, 11]. However, no known work has evaluated them in the autonomous driving domain.

End-to-end (E2E) autonomous driving is an approach to building systems where the transformation from sensor inputs to actuator control is learned directly with a single model using gradient-based learning. This is in contrast to a modular design, where there are either several machine-learning models doing separate tasks or some of the modules are hand-engineered.

According to [19], 1.35 million people die, and tens of millions are injured in car accidents worldwide every year. If you are between 5 and 29 years old, road traffic injury is the most likely way to die. Human error is the leading cause (around 94%) of car accidents [2]. Hence, a consistently safe solution for autonomous driving will have a chance to save the lives of up to a million people a year.

E2E learning is one of the most promising modern approaches to autonomous driving. If benefits of EBMs over standard models transfer to this domain, this might speed up the development of autonomous driving technology.

1.1 Contributions

The contributions of this thesis are three-fold:

- Show on a real car that EBMs can be at least or more effective at E2E autonomous driving as standard models.
- Introduce a temporal regularization loss term for smoothing out the energy landscape. It effectively reduces prediction differences over time, perceived as "jerkiness" of the steering wheel.

- Study the correlations between off- and on-policy metrics with a real car.

1.2 Outline

The thesis is organized as follows:

- Chapter 2 reviews definitions of behavioral cloning, energy-based models, and E2E autonomous driving.
- Chapter 3 describes the project's methods, such as the data and training pipeline, model architectures, metrics, and experiments.
- Chapter 4 presents the results of the experiments and findings.
- Chapter 5 discusses potential reasons and implications of the findings.
- Chapter 6 summarizes findings and contributions and outlines future work.

2 Background

This section briefly reviews the definitions of end-to-end autonomous driving, behavioral cloning, and energy-based models.

2.1 End-to-End Autonomous Driving

Most approaches to end-to-end autonomous driving rely on imitating human expert demonstrations via Imitation Learning (IL).

By far, Behavioral Cloning is the most researched Imitation Learning method [14].

2.1.1 Behavior Cloning

The problem formulation of IL is given below from [14]. Given a dataset $\mathbb{D}$ of expert state-action pairs (s, a) produced by an expert policy π^*, the goal of imitation learning is to train a policy $\pi_\theta(s)$ that maps any state s to a corresponding action a as closely to the expert as possible:

$$\arg\min_{\theta} E_{s\sim P(s|\theta)} L(\pi^*(s), \pi_\theta(s)), \tag{1}$$

where $P(s|\theta)$ is the state distribution of the trained policy π_θ.

Behavioral cloning reduces the IL task to a supervised learning problem. The objective of BC is to treat each state-action pair as an independent and identically distributed (i.i.d) example and minimize the imitation loss for the trained policy:

$$\arg\min_{\theta} E_{(s,a^*)\sim P^*} L(a^*, \pi_\theta(s)), \tag{2}$$

where the encountered state distribution, $(s, a^*) \sim P^*$, is now given solely by the expert policy and is assumed to be drawn i.i.d.

Following the terminology of [11], a behavior cloning policy function π_θ is called "explicit" when a policy maps the input observations $\mathbf{o}$ directly to output actions $\mathbf{a} \in \mathcal{A}$. An example of an explicit policy is a feedforward model of the form:

$$\hat{\mathbf{a}} = F_\theta(\mathbf{o})$$

2.2 Energy-Based Models

Any function that measures a kind of "goodness" value between two sets of variables, where "good" matches between two sets have a low value of this measure, can be called an energy-based model [15].

Using notation, given an input X, an energy-based model must produce the value Y^*, chosen from a set $\mathcal{Y}$, for which $E(Y, X)$ is the smallest:

$$Y^* = \arg\min_{Y \in \mathcal{Y}} E(Y, X) \tag{3}$$

Behavioral cloning policies are called "implicit" whenever π_θ from Equation 2 is a composition of argmin with a continuous energy function E_θ, such that:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a} \in \mathcal{A}} E_\theta(\mathbf{o}, \mathbf{a}) \tag{4}$$

3 Methods

In this chapter, details of the project will be described, such as the training pipeline, model architecture, training, and inference algorithms; a simple approach for temporal regularization will be introduced, and an intuitive uncertainty metric for EBMs will be proposed. Finally, the experiments and main quantities used for performance estimation will be described.

3.1 Dataset and Training Pipeline

The dataset and the explicit model training pipeline were borrowed from [18]. The dataset consists of 540 km of human driving, mostly on very low-traffic gravel roads. The recordings that make up the dataset are broken into training (460 km) and evaluation (80 km), with the evaluation recordings used for off-policy metric calculation and early-stopping. The original dataset includes camera and lidar images, but only camera frames are used in our experiments. The frames are cropped to remove the car's hood and everything beyond the horizon and limit the view to 90 degrees of the front center. The resulting frames of shape 264x68x3 are then normalized and fed into the model. The target labels correspond to the steering wheel angles of the human drivers.

For ordinary training, a mini-batch is created from RGB frames and corresponding steering wheel angles sampled uniformly at random from all recordings. For experiments with temporal regularization, a different sampling approach is used, where sequences of frames are sampled instead. The sequence length used in the experiments was two frames. The sequence dimension was flattened such that a mini-batch of sequences became a mini-batch of frames.

3.2 Model Architecture, Training and Inference

Both the explicit baseline and the energy-based model use the same backbone network from [18], which is, in turn, a modification of the classic Pilotnet architecture from [6]. The EBM model modifies the backbone by taking candidate actions as additional input (see Figure 1).

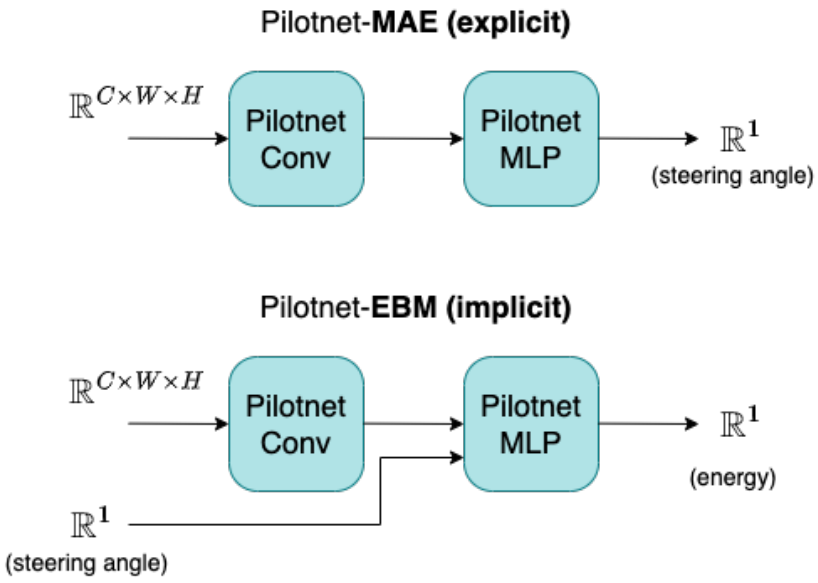

Figure 1. Model architectures for the explicit baseline and the EBM.

3.2.1 Training Procedure

The explicit model is trained to minimize the mean absolute error (MAE) between the predicted single output and the target steering value. The EBM model forward pass is as follows:

1. Take S scalar steering values (referred to as negatives) per minibatch item. Negatives are picked either uniformly at random or set to a static 1D grid.
2. Run the network forward pass to produce energy E for $S + 1$ (S negatives plus one target) steering values.
3. Compute the cross-entropy loss between softmax$(-1 * E)$ and the one-hot ground-truth vector.

Minimizing the cross-entropy of the *negative* of the energy has the effect of making the target steering value have a *low* value of "energy" and negatives have a *high* value of "energy". Taking the negative is not strictly necessary but is done for consistency with the EBM formulation.

The Adam [12] optimizer is used with default hyperparameters (learning rate $1 * 10^{-3}$, betas 0.9 and 0.999) and $1 * 10^{-2}$ weight decay [13]. Finally, early stopping is used on validation MAE with a patience of 10 epochs.

3.2.2 EBM Inference Procedure

Two different inference methods were tried: the Derivative-Free Optimization (DFO) algorithm [11] and a straightforward argmin method. DFO is an iterative sampling algorithm that improves upon the initial candidate actions. It is implemented as explained in [11]. The argmin method is defined simply as:

$$\arg\min_{A}(E_i), \tag{5}$$

where E_i is a shorthand for a vector of energies of the input and candidate actions at a time step i:

$$E_i = \left[\mathrm{E}(X_i, A_{i,1}), ..., \mathrm{E}(X_I, A_{i,n})\right] \tag{6}$$

where E is the energy function (an EBM), X_i is the input at timestep i, $A_i,_j$ is a candidate action j at timestep i, and n is the number of candidate actions.

Early experiments showed that DFO did not have better off-policy performance on our evaluation dataset than the simple argmin approach, so the latter was used in later experiments. The simple method was preferred as it slightly sped up the evaluation loop during training and was easier to implement. More importantly, the optimization loop of DFO relies on the Multinomial operation, which at the time of writing is not supported by TensorRT [3], the inference engine used in the car. Hence, most experiments with EBMs in the car were done with ONNX Runtime [4] as the inference engine. Explicit baseline models were verified to have the same performance across different engines.

Note, that in practice, energy values are computed in a single forward-pass for a mini-batch of inputs, where each input has a corresponding action vector. Inputs are repeated such that a mini-batch consists of *batchsize* $* S$ items, one per action in the mini-batch. Memory usage and visual processing time are minimized through late fusion (Figure 4b of [11]).

3.2.3 Car Hardware and Sofware Stack

The experiments were performed with Lexus RX 450h fitted with a PACMod v3 drive-by-wire system provided by AutonomouStuff. The following sensors were used: NovAtel PwrPak7D-E2 GNSS device, and a Sekonix SF3324 120-degree FOV camera. The camera works at 30 Hz. The car computer is equipped with a GeForce GTX 2080 GPU.

The car's end-to-end stack runs on ROS [17]. Before this project, ROS Melodic with Python 2 was used in the car. For this project, to support ONNX Runtime, Python 3 and ROS Noetic were installed. The Python 3 environment turned out to be slower than Python 2, dropping the effective model throughput from 30 Hz to less than 13 Hz. To cope with the slowdown, only the latest frames were used by the model. Equal performance between Python 2 and Python3 environments was empirically validated early on with the same models.

3.3 Evaluation Metrics

The explicit baseline and energy-based models were evaluated both off- and on-policy. It has been previously found that off-policy metrics do not necessarily predict true on-policy driving ability [7], hence the off-policy metrics are only computed to demonstrate the correlation between the two. The key reported off-policy metric is the validation set Mean Absolute Error (MAE) between the predicted commands and the human driver's true commands. Another metric reported, following multiple previous works [9, 10, 18], is the whiteness of the steering angle sequence, defined as:

$$W = \sqrt{\frac{1}{D}\sum_{i=1}^{D}\left(\frac{\delta P_i}{\delta t}\right)^2}, \tag{7}$$

where δP_i is the change in predicted steering angle, D is the size of the dataset, and δt is the time difference between decisions. Since our camera has FPS=30, $\delta t = 1/30 = 0.033$.

Whiteness measures mean smoothness of the generated command sequence and can be computed both off- and on-policy. $W_{effective}$ stands for the effective whiteness of the steering wheel position, $W_{command}$ is the whiteness of the model commands during an

on-policy evaluation, and $W_{off-policy}$ is the whiteness of the predictions computed on a driving recording.

It has been found by [18] that off-policy whiteness can correlate with on-policy performance hence the relationship between the metrics is evaluated in this work as well. Another potentially useful off-policy metric, EBM uncertainty, is calculated as described in Chapter 3.5.

The number of interventions during each on-policy evaluation was counted, and distance per intervention (DpI) was computed as the primary measure of a model's driving ability. Since the models were trained and evaluated only for road-following, interventions on intersections were subtracted from the total count. For safety reasons, the driver always took over the driving in case of oncoming traffic. Such interventions were similarly deducted from the total intervention count.

3.4 Temporal Regularization

The Results section shows that EBMs tend to have a higher whiteness than baseline explicit models. A higher steering wheel jerkiness may cause discomfort for the passengers, so it is worth trying to understand its cause and try to ameliorate it. A discussion of the possible cause for EBMs' higher whiteness is given in the Discussion section. Here an intuitive regularization term is proposed that can be added to the loss function to improve temporal smoothness.

To motivate the temporal regularization term, it is helpful to look at whiteness as a measure of sensitivity to small differences in the input space. If one could reduce the model's sensitivity to slight differences in the subsequent RGB frames, one would achieve temporally smoother predictions, i.e. lower whiteness. Hence, the regularization term should measure the prediction difference between subsequent frames. An obvious choice in the case of energy-based models is to measure the difference between the predicted energy of candidate actions, not the final output action. This will have the effect of smoothing the whole energy landscape. With these considerations in mind, the final regularization loss term is:

$$L_{temp} = \alpha D\Big(E_i, E_{i+1}\Big), \tag{8}$$

where α is the regularization strength, and D is some distance-like measure. Several measures have been experimented with: L1 and L2 norms, and the Earth Mover's Distance. The L2 norm was picked as the most consistent of them. Well-performing regularization strengths were found empirically and likely depend on the dataset and any model modifications that impact the training loss norm.

Since the energy function takes candidate actions as input, energy values at two consecutive timesteps will differ not only due to slight image differences but also due to different action inputs. A correct way to compute the regularization term would then be to use the same candidate actions for such pairs of consecutive frames. In the experiments, however, it was found that a simpler solution is just as effective — instead of sampling random candidate steering values for each mini-batch item, a static vector of linearly-spaced values was used. This change makes the model even simpler while maintaining the performance and offering a possibility to compare output not just at subsequent timesteps, but across time as well. When computing the regularization term, energy for the target steering value is not counted, since it is always unique for every frame.

3.5 Prediction Entropy as EBM Uncertainty

One of the key benefits of energy-based models is increased interpretability. In the case of end-to-end lateral control, instead of simply receiving one scalar steering angle value as produced by explicit models, one can evaluate the desirability of *all* potential angle values, which can be used to estimate how confident the model is in its predictions. If one imagines the energy landscape of a continuum of steering wheel angles, one would expect to see a steep valley around the target values when the model is confident in its prediction and more flat regions when it is less certain.

Visualizing the energy landscape across time against the footage from the car's camera (Figure 2) can be a useful tool for qualitatively assessing the model's reliability or predicting challenging situations before deployment. For a more scalable tool, a quantitative metric would be helpful. Here, a simple measure of uncertainty of an EBM

Figure 2. Visualization of the energy landscape across time on a recorded drive. The green pointer shows the predicted steering angle value at the current timestep. **Left**: the landscape has a clearly defined valley around 0 steering angle, representing high confidence. **Right**: the valley is about to disappear, replaced by a flatter region. Not shown here, the flat region, lasting for 7 seconds, was accompanied by the model drifting from the road, causing an intervention. The images were cropped arbitrarily and do not represent the model view.

is proposed:

$$U(E, D) = \frac{\sum_{i=1}^{|D|} \mathrm{H}\big(\mathrm{softmax}(-E_i)\big)}{|D| \log S}, \tag{9}$$

where E is the energy function (the EBM), D is the evaluation dataset, H is the entropy of a discrete probability distribution, and S is the number of candidate actions per frame that are part of the probability distribution. Dividing by $\log(S)$ normalizes the value between 0 and 1. Higher values mean higher uncertainty.

The entropy of the continuous probability distribution is approximated by turning it into a discrete probability distribution at fixed intervals and using its entropy instead. Thus, EBM uncertainty can only be computed when the inference method uses a constant steering angles grid.

3.6 On-Policy Evaluation Procedure

The on-policy evaluation was performed on a 4.3 km section of SS20/23 Elva track, driven in both directions (see Figure 8 for the route map). The speed was set to 80%

of the speed a human driver used in the same location and direction, extracted from a prior recording. Similar to prior work [18], setting the speed to 100% was attempted but felt too dangerous with weaker models. The testing was performed in the Summer, the same season as the training dataset, so the vegetation looked similar hence the evaluation season can be considered in-distribution.

The evaluation track is narrow, and driving off the road edge is hazardous for the car, so the safety driver was free to intervene whenever they perceived danger. Thus, an intervention is defined as a situation where the driver perceived an excessive threat to the car or the passengers. An intervention was triggered when the driver applied force to turn the steering wheel. If the model turned the steering wheel simultaneously and in the same direction as the safety driver, it would not cause an intervention since no force was applied.

4 Results

In this section, the real-world driving ability of energy-based and baseline models is presented and compared. Off-policy experiments are done to evaluate the helpfulness of derivative-free optimization for energy-based model inference and to assess the correlation between off-policy and on-policy metrics.

4.1 Choosing the EBM Inference Method

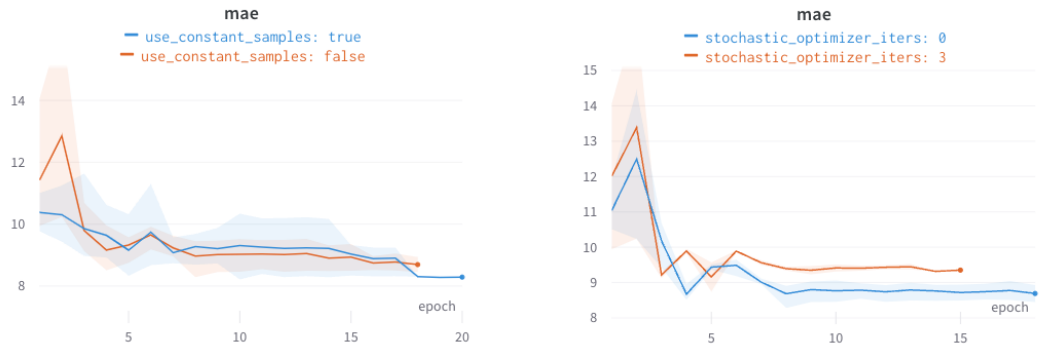

Figure 3. Average evaluation mean absolute error (MAE) curves. **Left**: EBMs with constant vs random steering angle samples. **Right**: EBMs using random action sampling and DFO with 3 iterations vs 0 iterations (i.e. simple argmin).

Two key axes of variations in the EBM inference methods are implemented in this thesis: how the action samples are drawn (uniformly at random or from a constant grid) and how the final steering value prediction is chosen (via a simple argmin or with DFO iterations). Off-policy experiments were used to choose a single configuration due to the prohibitive amount of time taken by on-policy evaluations.

First, it was expected that using a constant candidate action grid would have lower performance than using random samples for every mini-batch item since the randomization of actions is hypothesized to induce a kind of implicit regularization [11]. However, the off-policy experiments did not show a clear difference in performance between the two approaches (Figure 3 Left). Since no difference was seen, and temporal regularization

and EBM uncertainty require identical action inputs between steps, the constant candidate action grid was used in training and evaluating EBMs in on-policy experiments.

Derivative-free optimization (DFO) should improve on the initial candidate action samples. Since DFO can suggest actions that are not part of the input, a model trained with constant candidate actions would have to evaluate actions that fall between the constant grid values. To eliminate this as a possible confounder, the usefulness of DFO was evaluated on models with random candidate action sampling. Figure 3 (Right) shows that DFO did not give any performance boost over the simple argmin approach — on the contrary, it showed worse results on the evaluation loss metric. Hence, DFO is not used in all further EBM experiments, and a possible explanation for this is given in the discussion.

4.2 EBM vs Baseline Models Driving Ability

4.2.1 Generalization Ability

This project aimed to empirically assess the suitability of energy-based models as end-to-end learners in the autonomous driving domain. The ultimate goal of end-to-end driving is to have models that can generalize to unseen roads and conditions. Hence the evaluation was done on a track that was not in the training dataset but is similar.

Three different model types were evaluated on the real car:

1. EBM
2. EBM with temporal regularization
3. Baseline explicit MAE model

Three different initializations were tested per each model type, to estimate the stability of the results. The regularized models were each trained with a different temporal regularization strength parameter. Only one full forward and backward drive on the track was done per model due to the complexity of real-world evaluation. The evaluations were done by the same safety driver, on two different days, in sunny, clear-sky weather, at the same time of day. Sunny weather is not ideal for these models, because they were

Table 1. Results of on-policy evaluations.

Model (session)	Distance	Interventions	DpI	$W_{command}$	$W_{effective}$
EBM v1	8530.92m	12	710.91m	203.29°/s	43.82°/s
EBM v2	8565.79m	5	1713.16m	244.25°/s	35.15°/s
EBM v3	8552.43m	5	1710.49m	213.41°/s	38.06°/s
EBM Reg v1 ($\alpha = 30$)	8464.75m	9	940.53m	59.50°/s	31.62°/s
EBM Reg v2 ($\alpha = 10$)	8486.88m	7	1212.41m	66.36°/s	28.67°/s
EBM Reg v3 ($\alpha = 100$)	8292.14m	18	460.67m	57.78°/s	29.45°/s
MAE v1	8570.36m	5	1714.07m	96.05°/s	35.65°/s
MAE v2	8510.03m	8	1063.75m	82.51°/s	39.76°/s
MAE v3	8392.65m	16	524.54m	79.54°/s	46.10°/s

trained on the dataset with a mostly cloudy sky, so the weather could be considered to be slightly out-of-distribution for all tested models. This makes the results not directly comparable to previous evaluations on this track [18], but the relative performance of models trained in this project should still be comparable with each other. The results for these evaluations are presented in Table 1.

The results indicate that, on a novel track and in out-of-distribution weather, the performance of unregularized energy-based models is similar to or better than the standard explicit models. Regularized models seem clearly overregularized, with stronger temporal regularization strengths consistently causing worse performance.

4.2.2 Stability at Intersections

As noticed by passengers during the evaluation, baseline explicit models regularly exhibited swerving behavior at intersections when more than one route was possible (see an example in Figure 4). Stronger models would often recover quickly from the sudden turn, but for weaker models, this caused interventions. Even if no intervention occurred because of quick recovery, the safety driver considered the swerves unsafe.

In contrast, even unregularized energy-based models handled the same locations smoothly. As can be seen in Figure 4 (c), the EBM's energy landscape shows some

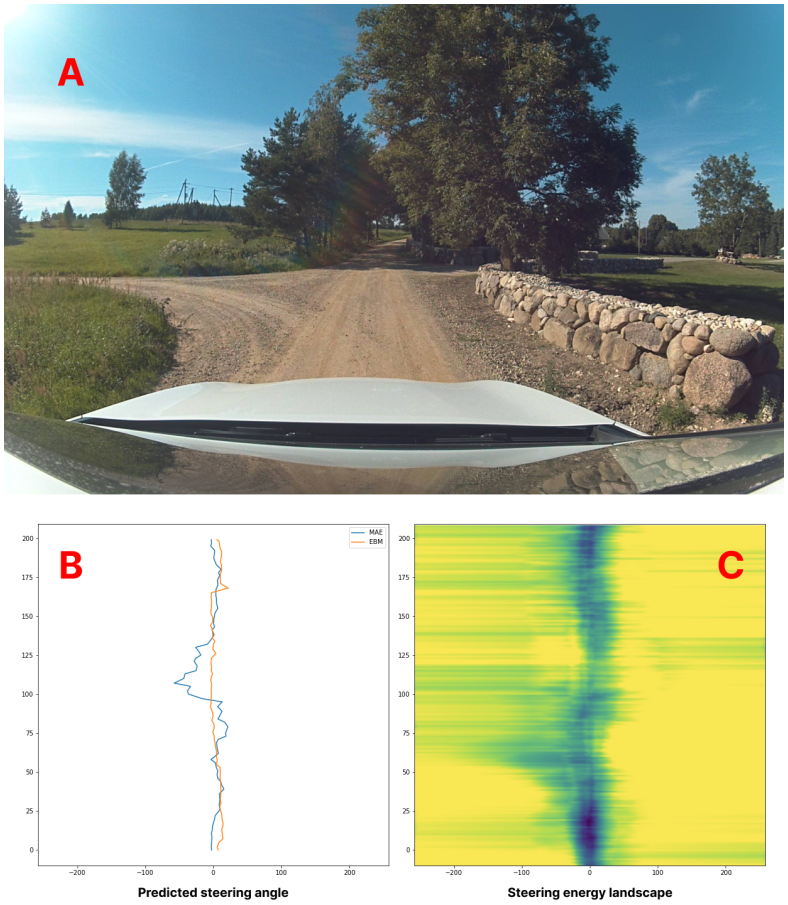

Figure 4. Example of swerving behavior by MAE models, not exhibited by an EBM

probability weight assigned to the turn, but the energy value closer to zero steering value is lower, and so the argmin action selection allows the model to drive smoothly straight (Figure 4 b).

On the other hand, the explicit MAE model, which could be thought of as modeling a Gaussian distribution, is incapable of learning a multimodal function, so its mean prediction is shifted toward the turn, resulting in swerving behavior. This is a practical example of EBMs' better ability to handle multimodalities, which are abundant in real-world driving. It may be that evaluating the models on more complex roads with a higher number of intersections could more strongly highlight the performance superiority of EBMs.

4.3 Temporal Regularization Effectiveness

Results show that unregularized EBMs have higher prediction whiteness than baseline explicit models (Table 1). In the case of the car used in the evaluation, this has not led to increased *effective* whiteness relative to the baseline models however it might still be important to have a solution for jerky predictions for cars where prediction whiteness transfers more readily to steering wheel's effective whiteness.

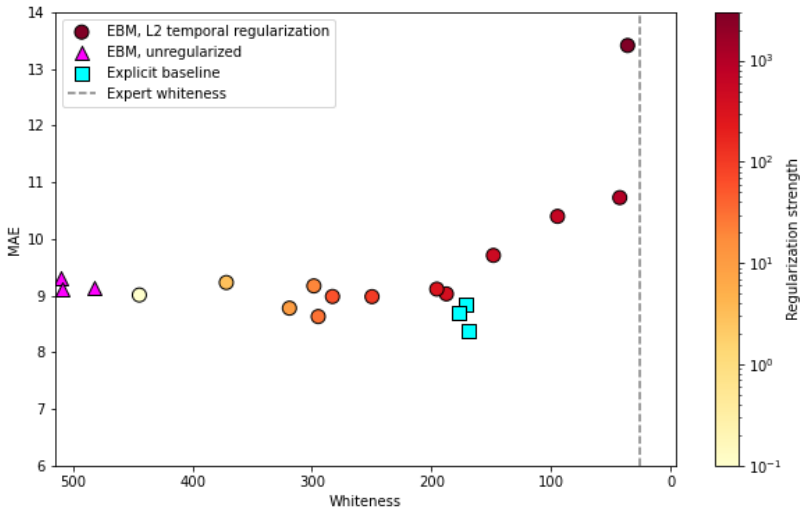

Figure 5. Temporal regularization strength effect on offline Whiteness and MAE

Table 1 shows that regularization significantly impacted prediction and effective whiteness but also negatively impacted performance. In off-policy experiments, there is a range of regularization strengths that reduce offline whiteness without affecting MAE (Figure 5), so it may be possible that there is a similar range of regularization strengths that would not affect real-world performance while decreasing prediction whiteness. Exploring whether command whiteness transfers to high effective whiteness with other car hardware and software stacks, as well as finding out a more suitable range of regularization values is left for future work.

4.4 Metric Correlation Study

Considering the complexity of the real-world evaluation of end-to-end models, finding an offline metric that is a good predictor of on-policy performance would be a major contribution to the field's development. Unfortunately, the experiments in this project have not shown any statistically significant correlations between offline and online metrics (Table 2).

Table 2. Pearson Correlations of DpI and other metrics (p-values shown in parentheses, none are significant at $\alpha = 0.05$)

Value	$W_{command}$	$W_{effective}$	$W_{off-policy}$	$MAE_{offline}$	U_{EBM}
Pearson R	0.5053	-0.1813	0.2085	-0.0536	-0.6696
	(0.1652)	(0.6406)	(0.5904)	(0.8911)	(0.1457)

Separate group correlations for each model class (EBM vs MAE), model type (EBM, EBM regularized, and MAE), and regularization presence were also computed but similarly did not result in any statistically significant correlations.

The p-values for some metrics hypothesized to correlate with driving performance are close enough to $\alpha = 0.05$, so increasing the number of evaluations might produce significant results. Ironically, and somewhat obviously, to allow others not to do many on-policy evaluations, one has to do a lot of on-policy evaluations.

5 Discussion

This section proposes explanations of the results and discusses the difference of energy-based models from a similar model class.

5.1 Difference of EBMs from Discretized Softmax

Discovering that more complex variants of EBMs did not have better performance in offline experiments led to a choice of a much simpler training and inference procedure. It is easy to notice that the resulting model is very similar to a simple softmax classification model, applied to a discretization of a continuous prediction problem. Being an implicit function, a simple classification model, trained with a cross-entropy loss, should be able to handle multi-modalities and discontinuities just like EBMs.

The only difference between the two models is that in an EBM, the candidate actions are represented explicitly and are fed as input to the model. Such representation has some benefits over the conventional classification, such as suffering less from catastrophic forgetting [16]. In addition to better multimodality handling, discussed in the Results, this may be another reason for EBMs' better performance in the evaluations.

5.2 Why Do EBMs Have Higher Whiteness?

Unregularized EBMs have been shown to have higher prediction whiteness than the baseline explicit models (Table 1). This subsection includes speculation about a possible reason for this.

First, the primary sources of whiteness variation among energy-based models should be considered. Two parameters were found to positively correlate with whiteness: the sampling bounds size of the range that training negatives are sampled from and the total number of negatives (Figures 6 and 7).

Consider the following in the context of EBMs trained with constant negatives, although this also applies to random sampling. If we were discretizing a continuous steering value space, larger bounds with all else same and more negatives with all else same could be considered analogous to (a) larger approximation bins and (b) more

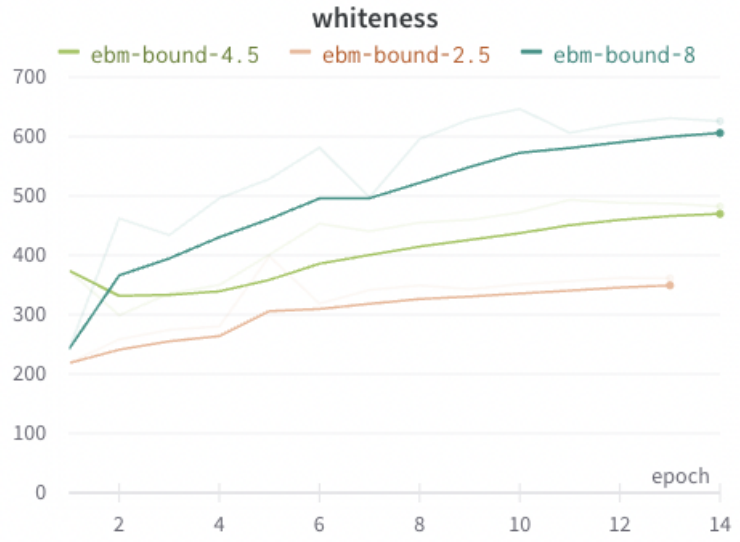

Figure 6. Whiteness grows consistently with wider negatives sampling bounds when keeping the number of negative counter-examples the same

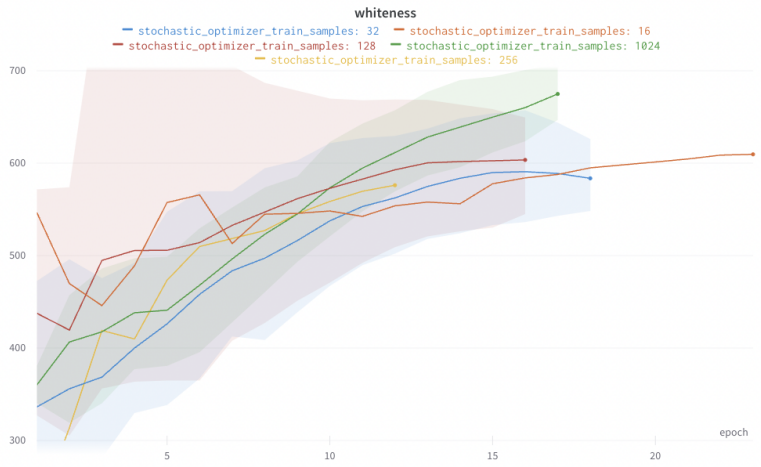

Figure 7. Whiteness tends to increase with a larger number of negatives ("stochastic_optimizer_train_samples") while keeping the number of inference samples and bounds size the same

approximation bins.

Thinking this way, it is natural that larger bins increase whiteness since sharp proba-

bility densities cannot be approximated that well.

On the other hand, a higher number of approximation bins causing higher whiteness seems less intuitive at first glance since *more* bins also mean *smaller* bins. Smaller bins approximate things better, and so should *decrease* whiteness. However, more bins also make training harder (final training loss indeed has a 0.95 Pearson correlation coefficient with the number of negatives) by reducing the number of times each bin is a target during cross-entropy loss minimization, hence making the energy landscape rougher.

The latter effect is likely partially offset by the reverse relationship between the bin size and quantity but should be visible when the two variables grow together.

If the above is true, then the single factor that explains the most of whiteness variance is simply the approximation precision of the true steering angle probability distribution. In that case, it may not be surprising that the EBMs trained with the cross-entropy loss produce a worse approximation than explicit models trained with MAE because the cross-entropy loss is not very efficient: e.g. there is no concept of label proximity, so when a learning step increases the log probability of the target steering value, it equally decreases the log probability of all other values, even if some are very close to the target value. An evaluation of EBMs with alternative, more efficient loss functions is left for future work.

5.3 Why Is DFO Not Helping?

Offline experiments did not observe DFO improve upon the initial candidate action samples. Instead, applying DFO decreased the performance. This subsection shares a speculative explanation of why this happened.

Considering that the DFO sampling is guided by the EBM, DFO consistently producing worse actions means one thing: the EBM does not approximate the energy landscape well. If it was, worse actions would not be selected by DFO, and the performance difference would not be as significant as seen in Figure 3.

Since EBMs with DFO were successfully used in previous work [11] in a different domain, it is natural to suggest that something about the task in this project makes it particularly hard to learn well.

One such thing could be the combination of label imbalance inherent to the automobile steering prediction (dense around zero and sparse further away) with uniform or linearly-spaced negatives sampling.

In that case, non-linear action sampling that follows the true action distribution and possibly a proximity-aware loss function (e.g. the Earth mover's distance) are likely to improve training efficiency, reduce prediction whiteness, and even make DFO further improve performance. Developing such action sampling and experimenting with proximity-aware loss functions is left for future work.

6 Conclusion and Future Work

The objective of this thesis was to evaluate the suitability of energy-based models for end-to-end autonomous driving. During the project, several variants of energy-based and baseline models were implemented, trained, and evaluated on a real car on a rally track. Several offline experiments were conducted to choose a better EBM variant, validate a proposed regularization term and an uncertainty metric, and study offline and online metrics correlation.

The key result is that a naive variant of an energy-based model has similar or better performance at end-to-end driving than the baseline models. Apart from having a slightly higher average distance per intervention, EBMs behave more stable at intersections, which is explained by their ability to handle multimodalities.

The trained EBMs have a higher whiteness, i.e. temporal prediction jerkiness, than the baseline models. Applying the proposed temporal regularization term significantly improves the models' whiteness, albeit slightly reducing performance. A possible cause of the higher whiteness is given in the discussion.

Offline and online metric correlations did not turn out to be statistically significant, so no conclusions could be drawn about the proposed uncertainty measure, but it is suggested that increasing the number of real-world evaluations might produce statistically significant results.

Future work on EBMs for autonomous driving should include the study of non-linear (or non-uniform) action sampling methods and proximity-based loss functions. As argued in the discussion, these two changes might be able to speed up the training and reduce prediction whiteness without the need for temporal regularization.

On another front, EBMs should be evaluated in a more complex setting, with numerous intersections and other sources of multimodalities. It would be interesting to see EBMs compared directly against other models which support multimodality, such as mixture density networks [5], used by an open-source end-to-end ADAS system Openpilot [1].

References

[1] Github - commaai/openpilot: openpilot is an open source driver assistance system. openpilot performs the functions of automated lane centering and adaptive cruise control for over 150 supported car makes and models. — github.com. [Accessed 08-Aug-2022]. URL: https://github.com/commaai/openpilot.

[2] Critical reasons for crashes investigated in the national motor vehicle crash causation survey. 2018. URL: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812506.

[3] Nvidia tensorrt, 2022. URL: https://developer.nvidia.com/tensorrt.

[4] Onnx runtime, 2022. URL: https://onnxruntime.ai.

[5] Christopher M Bishop. Mixture density networks. 1994.

[6] Mariusz Bojarski, Chenyi Chen, Joyjit Daw, Alperen Değirmenci, Joya Deri, Bernhard Firner, Beat Flepp, Sachin Gogri, Jesse Hong, Lawrence Jackel, Zhenhua Jia, BJ Lee, Bo Liu, Fei Liu, Urs Muller, Samuel Payne, Nischal Kota Nagendra Prasad, Artem Provodin, John Roach, Timur Rvachov, Neha Tadimeti, Jesper van Engelen, Haiguang Wen, Eric Yang, and Zongyi Yang. The nvidia pilotnet experiments, 2020. URL: https://arxiv.org/abs/2010.08776, doi:10.48550/ARXIV.2010.08776.

[7] Felipe Codevilla, Antonio M. López, Vladlen Koltun, and Alexey Dosovitskiy. On offline evaluation of vision-based driving models, 2018. URL: https://arxiv.org/abs/1809.04843, doi:10.48550/ARXIV.1809.04843.

[8] Yuntian Deng, Anton Bakhtin, Myle Ott, Arthur Szlam, and Marc'Aurelio Ranzato. Residual energy-based models for text generation. 2020. URL: https://arxiv.org/abs/2004.11714, doi:10.48550/ARXIV.2004.11714.

[9] Hesham M. Eraqi, Mohamed N. Moustafa, and Jens Honer. End-to-end deep learning for steering autonomous vehicles considering temporal dependen-

cies, 2017. URL: https://arxiv.org/abs/1710.03804, doi:10.48550/ARXIV.1710.03804.

[10] Nelson Fernandez. Two-stream convolutional networks for end-to-end learning of self-driving cars. 2018. URL: https://arxiv.org/abs/1811.05785, doi:10.48550/ARXIV.1811.05785.

[11] Pete Florence, Corey Lynch, Andy Zeng, Oscar Ramirez, Ayzaan Wahid, Laura Downs, Adrian Wong, Johnny Lee, Igor Mordatch, and Jonathan Tompson. Implicit behavioral cloning, 2021. URL: https://arxiv.org/abs/2109.00137, doi:10.48550/ARXIV.2109.00137.

[12] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. URL: https://arxiv.org/abs/1412.6980, doi:10.48550/ARXIV.1412.6980.

[13] Anders Krogh and John A. Hertz. A simple weight decay can improve generalization. In *Proceedings of the 4th International Conference on Neural Information Processing Systems*, NIPS'91, page 950–957, San Francisco, CA, USA, 1991. Morgan Kaufmann Publishers Inc.

[14] Luc Le Mero, Dewei Yi, Mehrdad Dianati, and Alexandros Mouzakitis. A survey on imitation learning techniques for end-to-end autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–20, 2022. doi:10.1109/TITS.2022.3144867.

[15] Yann Lecun, Sumit Chopra, and Raia Hadsell. *A tutorial on energy-based learning*. 01 2006.

[16] Shuang Li, Yilun Du, Gido M. van de Ven, and Igor Mordatch. Energy-based models for continual learning, 2020. URL: https://arxiv.org/abs/2011.12216, doi:10.48550/ARXIV.2011.12216.

[17] Stanford Artificial Intelligence Laboratory et al. Robotic operating system. URL: https://www.ros.org.

[18] Ardi Tampuu, Romet Aidla, Jan Are van Gent, and Tambet Matiisen. Lidar-as-camera for end-to-end driving, 2022. URL: https://arxiv.org/abs/2206.15170, doi:10.48550/ARXIV.2206.15170.

[19] World Health Organization. *Global status report on road safety 2018*. World Health Organization, "Geneve, Switzerland", jan 2019. URL: https://www.who.int/publications/i/item/9789241565684.

Appendix

I. Evaluation Route

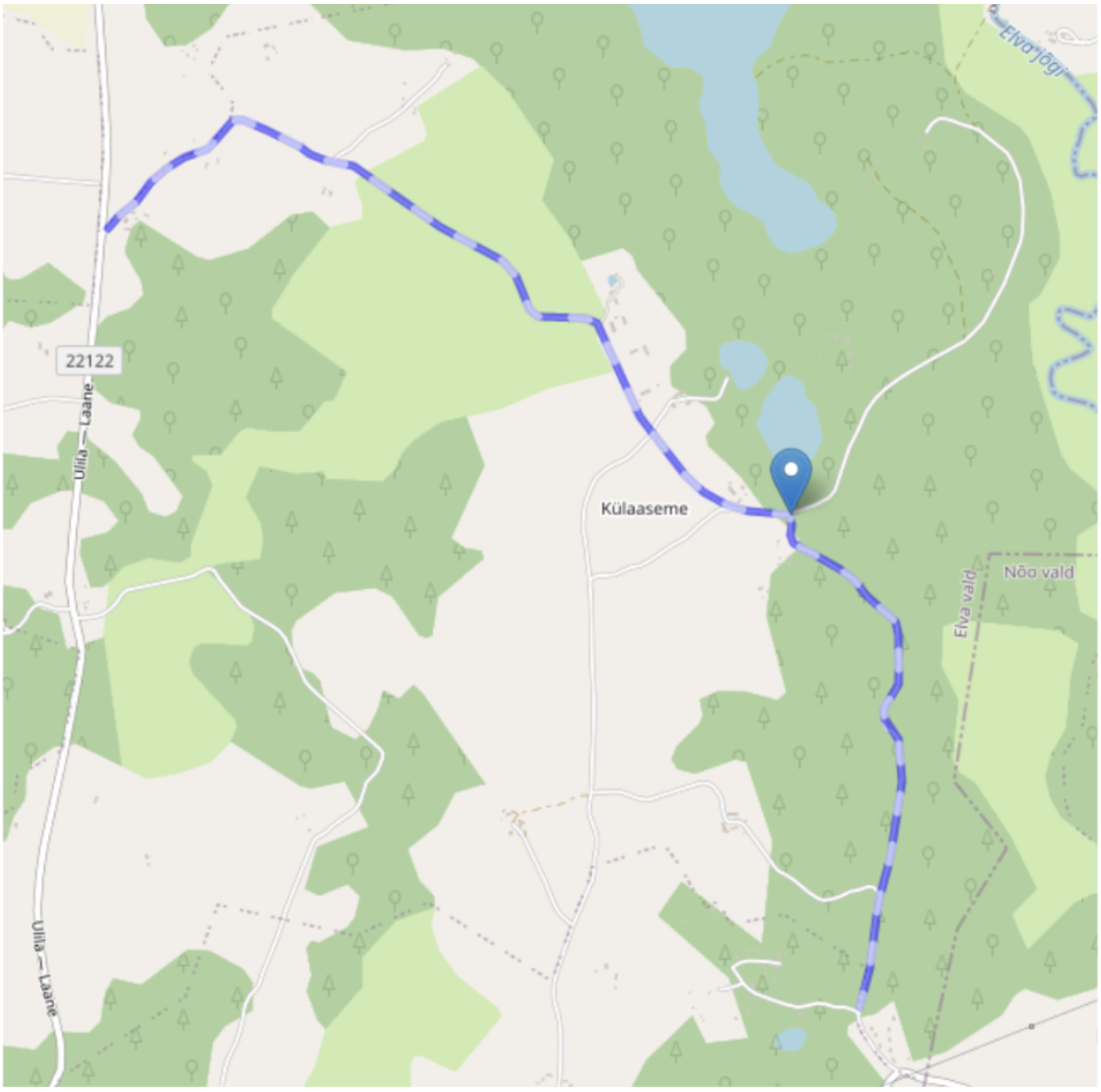

Figure 8. A 4.3 km section of the SS20/23 Elva track, used for the on-policy evaluations.

II. Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, **Mykyta Baliesnyi**,
*(*author's name)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

   Energy-Based Models for End-to-End Autonomous Driving,
   *(*title of thesis)

   supervised by Tambet Matiisen.
   *(*supervisor's name)

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Mykyta Baliesnyi
08/08/2022