ACHIEVING REALISM IN TRAFFIC SIMULATIONS: PERFORMANCE OF A COGNITIVE BEHAVIOR MODEL ON THE WAYMO OPEN SIM AGENT CHALLENGE

Lukas Brostek **Christian Rössert** **Johannes Drever** **Alexander Knorr**

cogniBIT GmbH
Agnes-Pockels-Bogen 1
80992 München, Germany
info@cognibit.de

June 3, 2024

ABSTRACT

Modeling human behavior in traffic is crucial for the development and testing of autonomous vehicles. Traditional rule-based models, while effective in ensuring collision avoidance and adherence to traffic rules, fail to capture the full spectrum of human behavior. Conversely, data-driven models utilizing machine learning techniques, such as neural networks, offer more realistic simulations but require vast amounts of data, especially for rare but critical situations like accidents. This paper introduces *cogniBOT*, a cognitive behavior model that integrates both rule-based and data-driven approaches by incorporating existing knowledge about human information processing. Our model, which uses significantly fewer parameters than purely data-driven models, is capable of adapting to both common and rare traffic situations. We demonstrate that *cogniBOT*, without being fitting on the provided training data set, achieves competitive performance in the Waymo Open Sim Agent Challenge, coming close to state-of-the-art data-driven models in terms of realism and interaction metrics. In contrast to its competitors, our general-purpose road user model has a broad range of applications while maintaining transparency and explainability.

1 Introduction

To date, two methods have primarily been used to model human behavior in traffic. Classical rule-based models simulate traffic behavior based on a combination of control algorithms for distance control and lane keeping. These relatively simple models focus on collision avoidance and adherence to traffic rules but do not cover typical human limitations and are unable to replicate the entire spectrum of human behavior in traffic.

A currently popular approach is to implement purely data-driven behavior models using machine learning techniques, such as artificial neural networks. These methods can reproduce traffic behavior in a much more realistic way. Due to the complexity of human behavior in traffic, a very large amount of data is required to train the networks. While large amounts of traffic data for standard driving situations are available, it is questionable whether sufficiently large amounts of data will be available for critical driving situations such as accidents and near-accidents, as these kinds of situations are extremely rare. However, valid simulation of critical driving situations is of the utmost relevance for safeguarding autonomous vehicles.

In recent years, a number of public challenges have been established in which simulated trajectories are compared with real trajectories to assess the fidelity or validity of human behavior models for the simulation of traffic. The two most prominent examples are the NuScenes Prediction Challenge (Caesar et al. 2020) and the Waymo Open Sim Agent Challenge (WOSAC, Montali et al. 2024).

While other successful models participating in these challenges are trained on specific large training datasets and have several million parameters, our model only uses generally available data and parameters from scientific publications, and therefore consists of only hundreds of parameters.

In this work, we show that our cognitive behavior model, *cogniBOT*, comes close to the successful data-driven models when accounting for map conversion and perception initialization errors.

2 Methods

We pursue a modeling approach that combines the advantages of rule-based and purely data-driven methods. The basic idea behind our approach is to incorporate existing knowledge about human information processing into a cognitive architecture. This architecture includes separate models for perceptive, cognitive, and motor sub-processes. The subdivision and usage of existing knowledge about sensorimotor processing steps allow us to drastically reduce the number of free model parameters, which also reduces the need for training data. This means that our behavior models can be adapted not only to normal traffic situations but also to the very rarely occurring events, such as critical driving situations.

Our general-purpose road user behavior models are based on the cogniBOT system architecture, a neuro-cognitive model of human information processing. The different sub-processes involved in the transformation of sensory information into situation-adapted actions are explicitly modeled. In the concrete application of modeling road user behavior, the following processes are of specific relevance. They can be structurally divided into three functional groups and are depicted in Figure 1.

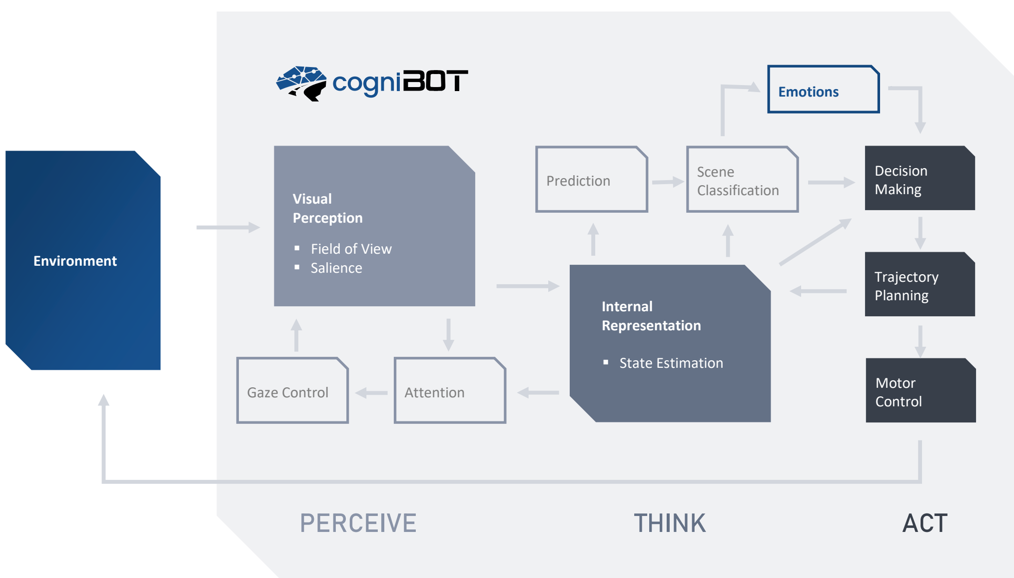

Figure 1: The cogniBOT system architecture

Visual perception describes the intake of visual information from the environment. The cogniBOT system architecture simulates relevant limitations of human road users, for example, a restricted field of vision, which is compensated by eye movements. The simulated eye movements are controlled by a complex attention process that takes into account both top-down signals, such as the currently intended action, and bottom-up signals, for instance, due to the recognition of other traffic participants in the peripheral field of vision.

The cogniBOT system architecture is a stochastic model that simulates processes probabilistic in nature as well. Limitations or inaccuracies in human perception are modeled by adding noise to the processed signal. This is also how, for instance, the acuity distribution in human vision, i.e., the decrease of visual acuity for peripheral vision, is modeled.

Another stage in cognitive processing where stochastic variation and 'errors' frequently occur are cognitive judgments and predictions of all kinds, such as when judging the suitability of a gap when performing a lane change maneuver.

The information recorded in the perception module is used to form an internal representation of the external world. In this process, an internal representation of the environment is formed from the visual situation, which also reflects only the objects actually recognized. Explicit modeling of this module allows the agent model to realistically describe characteristic misjudgments, such as the frequently occurring overestimation of the distance to vehicles in the rear-view mirror.

Just as the human driver anticipates and forms a prediction about the further development of a traffic situation, the cogniBOT architecture draws on previously identified information about the type, position, and speed of other road users, as well as the internal map of the road course to create a context-specific prediction from this information. In complex traffic scenarios, cogniBOT uses a prediction mechanism in which each simulated agent anticipates the behavior of all other road users based on its individual assessment of the traffic situation.

The prediction of the situation forms the further basis for the decision of the simulated agent for a concrete driving behavior. For this purpose, the cogniBOT architecture implements a cost function that allows the simulated agent to make a trade-off for each traffic situation between speedy progress, distance to other road users, and the resulting risk of an accident.

In order to be able to realistically simulate the entire range of human behavior, human emotions are also modeled by cogniBOT. This makes it possible to simulate a wide range of driver types, from relaxed 'cruisers' to aggressive 'speeders'.

The motor performance of individual road users is a significant cause of individual differences between road users, which an agent model must represent to show characteristic differences between, e.g., old and young road users.

The individual modules discussed above utilize a variety of technical approaches for implementation. These include, but are not limited to, algorithms derived from theoretical neuroscience and robotics and control engineering. For specific modules, such as scene classification or trajectory prediction, we employ classic machine learning approaches like neural networks (e.g., MLP, LSTM). Unlike purely data-driven approaches, we opt for small multi-parametric models with around 1,000 parameters for 'simplified' tasks, significantly reducing the complexity of the machine learning problem. These distinct classification or regression problems are trained using synthetically generated data.

The number of remaining free parameters of the overall model is in the order of 100 and thus far below that of purely data-based models. A number of these remaining parameters, especially those related to perceptive and motor control processes, correspond to physiological variables, such as the size of the field of vision or the dynamics of eye movements. Most of these parameters have been analyzed by scientific studies and can be determined from the academic literature. Parameters related to decision-making processes, such as those defining the distance-keeping behavior when driving, are determined using data from the traffic research literature (see Fig. 2). It is important to mention that the training dataset provided in the Waymo Open Sim Agent Challenge was not used to fit model parameters for the cogniBOT release tested here.

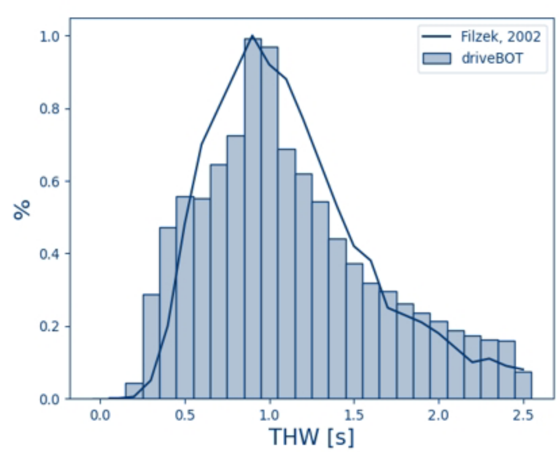

Figure 2: Distributions of time headway observed on German Autobahn (from Filzek 2003) and simulated by driveBOT v1.3 (the car driver model based on the cogniBOT system architecture)

3 Results

Figure 3 shows preliminary results of the cogniBOT road user model for a subset of 30,000 scenes randomly sampled from the 2024 Waymo Open Sim Agent Challenge (as of May 7th, 2024) in orange. For this sample, the cogniBOT behavior model achieved an average performance of 0.643 for the realism meta-metric. For each metric, there is a specific maximal achievable value, denoted as the 'logged oracle.' This maximum value is achieved if each trajectory realization corresponds to the actual recorded trajectory of a respective road user. The 'logged oracle' reaches an average value of 0.82 for the realism meta-metric. Remarkably, cogniBOT achieved values of 0.895 and 0.796 for the collision indication and time to collision metrics, respectively, reflecting the capability of simulating realistic interactions between road users.

The remaining collisions observed can mainly be divided into three categories. First, the dataset contains only incomplete or inconsistent information about the traffic light states for many intersection scenes. Estimating the correct traffic light states is still a challenge for cogniBOT in the tested version. Second, collisions involving pedestrians occur, which could be avoided if better indications of the pedestrians' navigation targets exist. Third, some scenarios start in the middle of active maneuvers where the lead-up for human perception accumulation is missing, leading to collisions.

With respect to the off-road indication metric, which is included in the metric with a share of 25%, there were a number of scenarios in which road users apparently left the drivable or passable area. We would like to point out that there are still some issues with the conversion of the map information provided at the current stage of the cogniBOT implementation. To record areas that can be driven and walked on, the map information is read in and converted into an internal map format. For this conversion, manual adjustments must be made for certain rarely occurring road geometries.

These shortcomings can be seen in the bimodal distribution of the meta-metric, which is caused by outlier scenarios with collision indication and off-road indication metrics smaller than 0.4. Removing these scenarios from the analysis (Fig. 3, green bars) leads to an overall meta-metric of 0.715, coming closer to the best data-driven model of 0.751 (as of May 28th, 2024).

4 Discussion

Our results demonstrate that the general-purpose road user model, cogniBOT, delivers competitive performance in terms of the realism meta score and is capable of realistically simulating standard traffic situations as they occur in the 2024 Waymo Open Sim Agent Challenge (WOSAC) dataset. Particularly in metrics related to interactions between road users, cogniBOT achieved good results.

The WOSAC realism meta metric includes various measures intended to quantify the degree of realism in a road user simulation. Metrics indicating the occurrence of collisions between road users or evaluating the distribution of time-to-collision are valuable evaluation parameters, as accidents are rare events in traffic. The off-road indication metric is also a useful evaluation criterion for passenger cars, as they are typically driven on roads. However, this criterion can be challenging to evaluate for pedestrians or cyclists, as both on- and off-road trajectories can be considered realistic.

Some of the evaluation criteria used in WOSAC assess the degree of reproduction of specific scenes rather than the general realism of the traffic simulation. For example, the distributions of angular speed and acceleration, linear speed and acceleration, and distance to the nearest object depend heavily on whether the navigation destinations of the simulated road users match those recorded in the scenes. The maximum value for these metrics is only achieved if each realization moves along identical routes. Estimating the desired speed of movement for a particular road user is also essential for achieving good evaluation results. In our view, a metric evaluating the realism of a road user simulation should rather focus on adherence to general parameters typically observed in reality, then on the specific reproduction of recorded trajectories.

Unlike all other road user modeling approaches competing in the 2024 Waymo Open Sim Agent Challenge, cogniBOT did not use the provided training dataset for fitting model parameters. Only minor adjustments were made to account for the perception of special road geometries and adaptation to local traffic rules. Hence, overfitting on the provided data material can be ruled out. Apart from the map conversion and initialization issues, the greatest potential for achieving a higher realism meta metric score lies in more accurately reproducing the recorded trajectories, particularly the navigation destinations and target speeds of individual road users. This, however, could be considered as a kind of overtraining.

Our general-purpose road user model cogniBOT scales across different contexts and scenarios. The high degree of generalizability allows the same model to be used for simulating standard traffic situations in various regions and

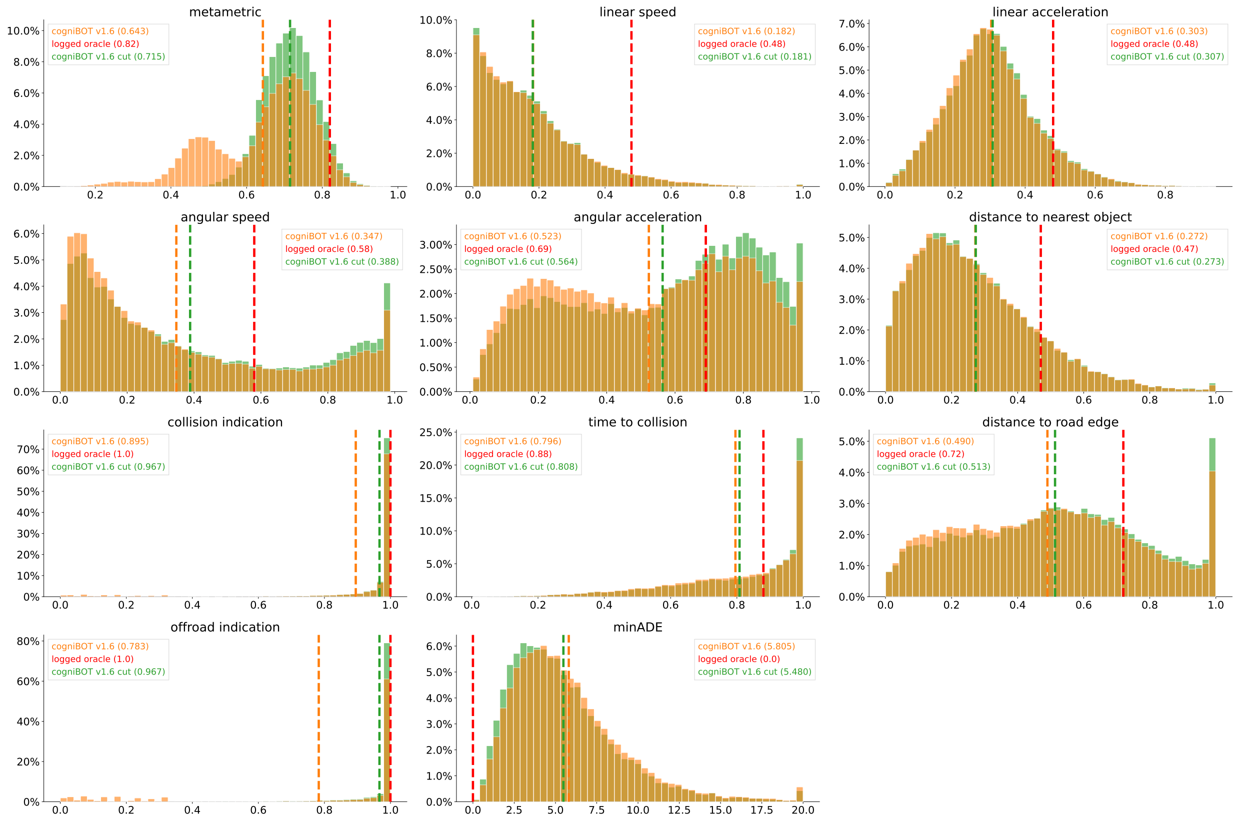

Figure 3: Evaluation results for the 2024 Waymo Open Sim Agent Challenge dataset. The figure shows the distributions of the evaluation results for cogniBOT for the individual metrics tested, as well as the realism meta-metric. While orange shows all results from the sub-sampled validation set, green depicts results when removing scenarios having map conversion errors (off-road indication < 0.4) and perception initialization errors (collision indication < 0.4). The maximum achievable average value for each metric, the 'logged oracle,' is also shown for comparison.

countries. The duration of the simulation scenes is not limited, enabling the simulation of more complex intersection negotiation problems. Based on the simulation of sensorimotor processes, cogniBOT can also realistically model behavior and reaction times in critical traffic situations, such as rear-end collisions. Further advantages of the cognitive modeling approach include inherent transparency and full explainability, the ability to simulate specific personalities (such as aggressive or prudent road users) and physiological conditions, and comparatively low computing requirements.

In conclusion, cogniBOT represents a significant advancement in road user behavior modeling by combining the strengths of rule-based and data-driven approaches. Its ability to realistically simulate traffic behavior with a reduced need for extensive training data makes it a valuable tool for the development and testing of autonomous vehicles and traffic systems. Future work on the WOSAC challenge will focus on refining the model to improve its handling of traffic light states and pedestrian navigation targets, as well as enhancing the map conversion process to further reduce errors and improve overall performance.

References

Caesar, Holger et al. (2020). "nuscenes: A multimodal dataset for autonomous driving". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11621–11631.

Filzek, Björn (2003). "Abstandsverhalten auf Autobahnen-Fahrer und ACC im Vergleich". In: *FORTSCHRITT-BERICHTE VDI. REIHE 12, VERKEHRSTECHNIK/FAHRZEUGTECHNIK* 536.

Montali, Nico et al. (2024). "The waymo open sim agents challenge". In: *Advances in Neural Information Processing Systems* 36.