



Article

Thrifty World Models for Applying Machine Learning in the Design of Complex Biosocial–Technical Systems

Stephen Fox ^{1,*} and Vitor Fortes Rey ^{2,3}

¹ VTT Technical Research Centre of Finland, 02150 Espoo, Finland

² DFKI German Research Center for Artificial Intelligence, 67663 Kaiserslautern, Germany; vitor.fortes_rey@dfki.de

³ Department of Computer Science, RPTU University Kaiserslautern-Landau, 67663 Kaiserslautern, Germany

* Correspondence: stephen.fox@vtt.fi

Abstract

Interactions between human behavior, legal regulations, and monitoring technology in road traffic systems provide an everyday example of complex biosocial–technical systems. In this paper, a study is reported that investigated the potential for a thrifty world model to predict consequences from choices about road traffic system design. Colloquially, the term thrifty means economical. In physics, the term thrifty is related to the principle of least action. Predictions were made with algebraic machine learning, which combines predefined embeddings with ongoing learning from data. The thrifty world model comprises three categories that encompass a total of only eight system design choice options. Results indicate that the thrifty world model is sufficient to encompass biosocial–technical complexity in predictions of where and when it is most likely that accidents will occur. Overall, it is argued that thrifty world models can provide a practical alternative to large photo-realistic world models, which can contribute to explainable artificial intelligence (AI) and to frugal AI.

Keywords: algebraic machine learning; behavioral ethics; biosocial–technical systems; explainable AI; frugal AI; sensory ecology; systems design; thrifty world models



Academic Editor: Jianlong Zhou

Received: 11 June 2025

Revised: 7 August 2025

Accepted: 8 August 2025

Published: 13 August 2025

Citation: Fox, S.; Fortes Rey, V. Thrifty World Models for Applying Machine Learning in the Design of Complex Biosocial–Technical Systems. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 83. <https://doi.org/10.3390/make7030083>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The construct of world models has origins in theoretical psychology, control theory, systems science, and computational neuroscience [1–6]. Human world models are mental models of the self in the World that can facilitate survival in the World. Human world models can generate inferences about what actions to take to change the World so it is better aligned with the world model, and about what updates are needed to the world model so it is more aligned with the World. In other words, human world models can enable active inference for survival in changing environments [6]. However, a human world model does not encompass the whole World. Rather, a human world model encompasses those aspects of the World that can affect survival.

In 2025, large machine learning world models are being introduced, which can include photo-realistic environments [7,8]. It is intended by their developers that some AI world models incorporate and understand physics [9]. However, AI world models are not consistent with the fundamentals of physics unless they are consistent with nature being thrifty in all its actions [10]. In particular, complex living systems evolve towards dimensional reduction [11], and there is natural selection for least action [12], which is a fundamental principle of physics [13,14].

In this paper, a study is reported that had the objective to investigate the potential for a thrifty world model to predict consequences from choices about road traffic system design. In particular, a world model comprising only three categories, which encompass eight design choices for road traffic systems. In the next section, Background, road traffic systems are explained as complex biosocial–technical systems. In Section 3, Model Structure, the thrifty world model for road traffic systems is explained. In Section 4, Model Implementation, the implementation of the thrifty world model with algebraic machine learning [15,16] is explained. In Section 5, Results, the results are described. In Section 6, Discussion, broader implications of the study are discussed. Overall, the reported study provides contributions to new directions for artificial world models [7–9], explainable AI [17], and frugal AI [18].

2. Background: Road Systems as Complex Biosocial–Technical Systems

Interactions between human behavior, legal regulations, and monitoring technology in road traffic systems provide an everyday example of complex biosocial–technical systems. A biosocial–technical system [19,20] is a framing that encompasses interrelationships between biological, social, and technological variables in complex phenomena. Consider, for example, some biological issues in the social practice of shopping from home enabled by Web-based technologies. More home deliveries can lead to delivery drivers sitting in vehicles for long hours as they try to earn enough to survive [21]. This can lead to negative biological consequences as sedentarism is related to multiple illnesses [22]. Moreover, in addition to the long-term personal health risks for delivery drivers, there could be immediate risks for the safety of drivers and other road users because psychological fatigue and physical fatigue can undermine self-regulatory control [23,24]. For example, a driver could have reduced self-regulatory control due to psychological and physical fatigue caused by many hours of driving that have involved having to stop at many traffic lights. If the driver is under time pressure [25] due to being delayed and under organizational pressure [26] to deliver quickly, the driver may be more likely not to slow down and stop when traffic lights are changing from amber to red.

Furthermore, gig-economy delivery drivers can have their payments for each delivery reduced if deliveries are not made on time [27]. Thus, time pressure and organizational pressure can take place in the context of gig-economy delivery drivers' survival being precarious [28]. Accordingly, there can be potential for behavioral ethics [29–31] to override virtue ethics [32], deontological ethics [33], and consequentialist ethics [34]. In particular, the sanctity of human life is an important belief in many cultures [35], which is expressed throughout the world in legal regulations that are intended to protect human life, such as road traffic speed limits [36]. From a virtue ethics perspective, people may believe that it is virtuous to protect human life by adhering to speed limits. From a deontological ethics perspective, people may believe that rules to protect human life, such as speed limits, are good and should be followed. From a consequentialist ethics perspective, people may believe that it is better for the journeys of many road travelers to be slowed rather than the lives of even a few people be endangered. Nonetheless, the behavior of drivers is not always consistent with virtue ethics, deontological ethics, or consequentialist ethics. Instead, people sometimes drive faster than speed limits, and gig-economy delivery drivers are under particular pressure to drive fast because of the general income precariousness and because of delivery-specific payment reductions [27,28].

Figure 1 provides a summary of three variables that can interact to influence whether a person's actions are in accordance with the normative ethics of what should be done, such as adhering to speed limits, or are in accordance with behavioral ethics, such as having to make many on time deliveries as quickly as possible in order to earn enough money to be

able to survive. In simple terms, whether a driver's actions are focused on the sanctity of human life in general [35] or are focused on the sanctity of the driver's own life [37].

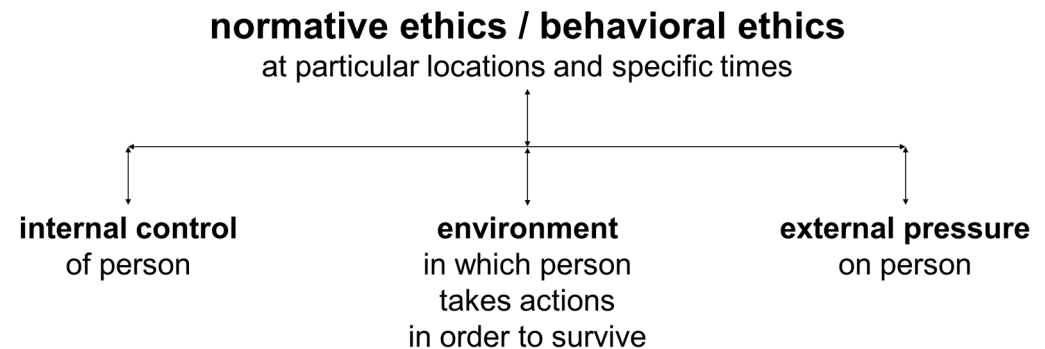


Figure 1. Behavioral ethics variables: internal control, external pressure, and survival environment.

The three variables are a person's internal control, the external pressure on a person, and the environment in which the person takes action in order to survive. Internal control, e.g., self-regulatory control, can be subject to external pressure such as time pressure and organizational pressure [25,26]. As illustrated in Figure 2, there can be changing balances between moral motivation enabled by internal control and ethical temptation due to external pressure.

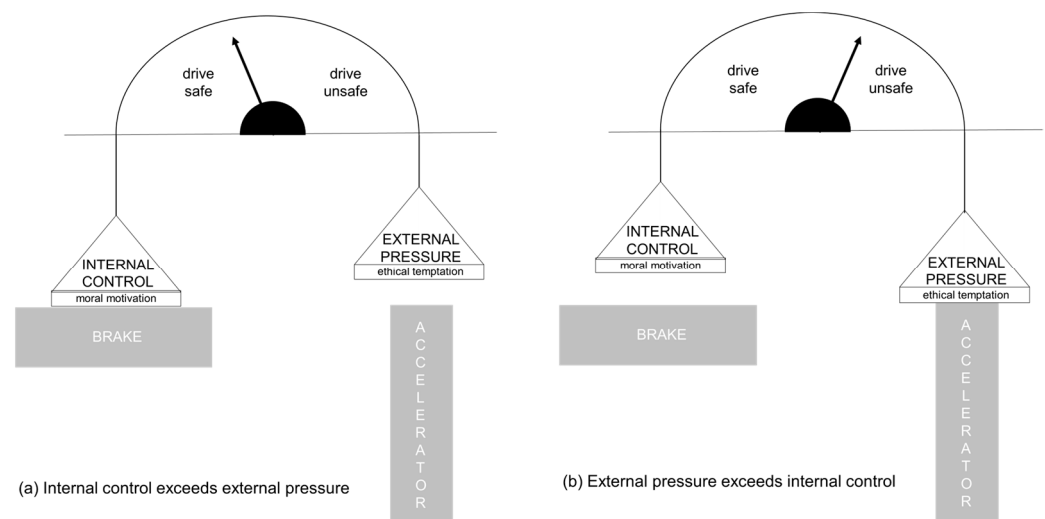


Figure 2. Changing balance between internal control and external pressure when driving.

However, environments can be engineered to reduce dependence on internal control. For example, narrow curved raised road surfaces, called speed bumps, can be constructed across roads. Speed bumps can slow driving speeds by causing vertical deflection of vehicles. The more the height of speed bumps is increased, the more the speed of vehicles is reduced [38]. However, as well as reducing vehicle speeds, speed bumps can slow the response time of emergency vehicles, may divert vehicles to nearby residential streets, and may increase noise and pollution for residents living immediately adjacent to the speed bumps [39]. Hence, speed bumps cannot be built along every road.

Overall, road traffic systems can be considered to be complex biosocial–technical systems, in which human internal control to adhere to social rules, such as speed limits, can be mediated through external pressures, such as delivery payment deductions, and technologies in the environment such as speed bumps. Interactions between biological, social, and technological variables can be difficult to predict; for example, because speed

reduction technologies can be situated in different locations that may or may not be on a particular delivery route.

3. Model Structure

Large world models that have photo-realistic environments are not thrifty enough to be aligned with physics [10–14] or with the urgent need for more frugal AI [18]. Figure 3 shows the thrifty world model structure, comprising three categories, including eight system design choice questions. Like a human world model, the thrifty world model does not encompass the whole World. Rather, it encompasses only those aspects of the World that affect survival. In this case, it encompasses aspects that affect the survival of the jobs of the local road traffic personnel who could be held accountable if their design of the local road traffic system increases road traffic accidents. The system design choice made for each of the eight questions is a prior belief about what action selection in the World can change road traffic risks. Together, chosen prior beliefs comprise the local road traffic authority's prior policy for road traffic systems safety.

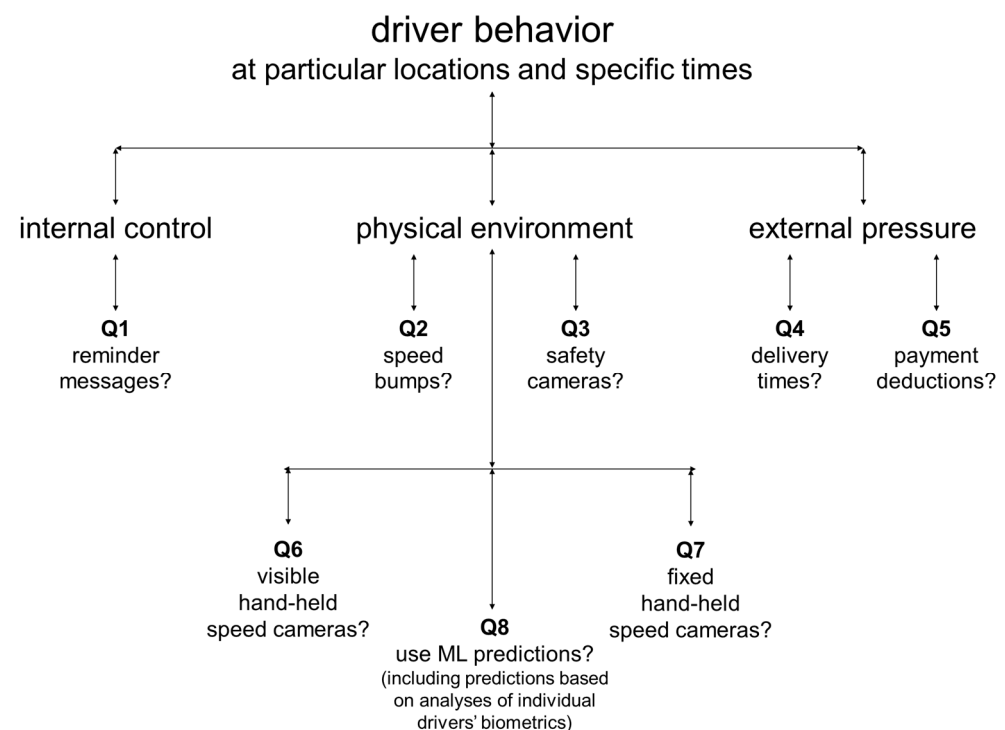


Figure 3. Thrifty world model comprising three categories, including eight questions.

The first system design question (Q1) concerns internal control and is as follows: Should all delivery drivers be sent automated messages to remind them about the need to prioritize safe driving, or is each individual delivery driver solely responsible for always driving safely?

The second question (Q2) and the third question (Q3) are concerned with the physical environment. Q2 is as follows: The local government road department has enough money to construct speed bumps in only one location. Should the speed bumps be constructed close to a kindergarten in the city or close to an elderly persons' home in the countryside? Q3 is as follows: The local government road department has enough money to install and operate one speed camera [40]. Should the road speed camera be set up close to a kindergarten in the city or close to an elderly persons' home in the countryside?

The fourth question (Q4) and the fifth question (Q5) are concerned with external pressure. Q4 is as follows: Question 4: How many minutes should delivery drivers be

allocated to make deliveries? The number of minutes indicated by a publicly available journey planner app or the number of minutes indicated by a publicly available journey planner app minus 10 percent, because delivery drivers are professional drivers? Q5 is as follows: How much money should delivery drivers be deducted per minute for late deliveries? Should they not have deductions from their payments if they deliver late, or should two euros be deducted for each minute that a delivery is late?

The sixth (Q6), seventh (Q7), and eighth (Q8) questions are concerned with more sophisticated design of the physical environment. Q6 is as follows: Should the local government road traffic department deploy local officials with hand-held speed cameras at roadside locations, where they can be seen from a far distance by drivers, or where they cannot be seen from a far distance by drivers? Q7 is as follows: Should the local government road traffic department deploy local officials with hand-held speed cameras at roadside locations, which are always the same locations, or which are at locations that often change? Q8 is as follows: Should the local government road traffic department deploy local officials with hand-held speed cameras at roadside locations who can access machine learning predictions of where road traffic regulations violations are most likely to happen, or who can access machine learning predictions of which delivery drivers are most likely to violate road traffic regulations?

The thrifty model comprises only three categories, internal control, external pressure, and physical environment, which include only eight system design choice options. Nonetheless, the thrifty model is sufficient for the definition of prior beliefs that together provide a prior policy for addressing complex biosocial–technical interactions in road traffic systems. As explained in the next section, these interactions were explored through the implementation of the thrifty world model using machine learning.

4. Model Implementation

Figure 4 below provides a summary of the implementation with algebraic machine learning [15,16] of the thrifty model described above. This is a classification model for predicting driver behavior, in particular, to predict where accidents are most likely to happen, so that the road traffic system can be redesigned to reduce accidents. For example, redesigns involving the repositioning of speed cameras to where accidents are most likely to happen in order to slow driving speeds and so reduce the number of accidents that actually happen.

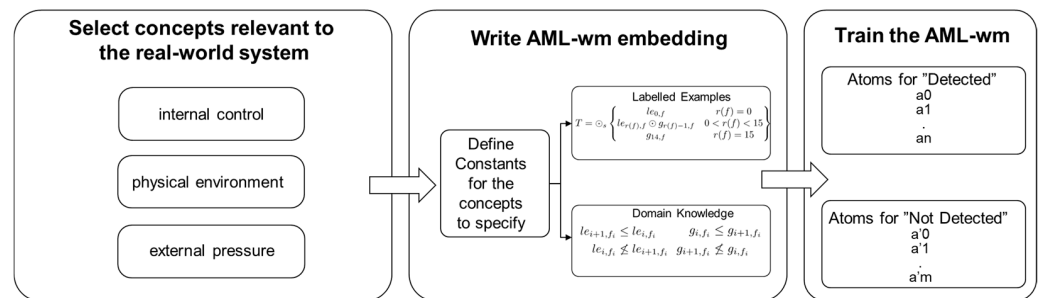


Figure 4. Implementation of the thrifty world model with algebraic machine learning.

Training of algebraic machine learning (AML) models is explained in detail in [15]. The diagram in Figure 4 summarizes the three stages of model formulation. First, select concepts relevant to the real-world system. The main concepts are internal control, physical environment, and external pressure. The second stage is writing the embedding by defining constants for the main concepts. This is performed through labeled examples in datasets, which, in this case, are datasets for driver fatigue, and through the definition of domain

knowledge, which, in this case, encompasses external pressure and physical environment. The third step is the training of the machine learning model.

The machine learning model is implemented with AML. For the embedding, using each feature f_i consisting of n_{f_i} levels of intensity $2 \times (n_{f_i} - 1)$ constants associated with the feature are as follows:

$$le_{0,f_i}, le_{1,f_i}, \dots, le_{n_{f_i}-1,f_i} \text{ and } g_{0,f_i}, g_{1,f_i}, \dots, g_{n_{f_i}-1,f_i}$$

The formal knowledge we want to express is the notion of intensity, which we represent as ordered sets:

$$\begin{aligned} le_{i+1,f_i} &\leq le_{i,f_i} \\ le_{i,f_i} &\not\leq le_{i+1,f_i} \\ g_{i,f_i} &\leq g_{i+1,f_i} \\ g_{i+1,f_i} &\not\leq g_{i,f_i} \end{aligned}$$

These form one ascending chain and one descending chain for each feature. The goal is to embed the discretized feature readings, which correspond to numerical values and are therefore represented as ordered sets in a discrete system. Every training example can be represented as a set of features representing feature values and one output value representing whether or not an accident occurred in the next fifteen minutes. For every training example, based on its $v(f_i)$ representing its discretized value of feature f_i , we create a term T representing the problem input as follows:

$$T = \odot_{f_i} \left\{ \begin{array}{l} le_{0,f_i} \odot v(f_i) = 0 \\ le_{v(f_i),f_i} \odot g_{v(f_i)-1,f_i} \odot 0 < v(f_i) < (n_{f_i} - 1) \\ g_{n_{f_i}-1,f_i} \odot v(f_i) = (n_{f_i} - 1) \end{array} \right\}$$

The training example is then embedded with a positive relation to either a constant O_0 (no accident) or a constant O_1 (accident). For the training label k , we have

$$\begin{aligned} O_k &\leq T \text{ Inclusion} \\ O_{k' \neq k} &\not\leq T \text{ Exclusion} \end{aligned}$$

The implemented world model was developed with historic accident data, which is modified in accordance with the set of design option choices. The resulting simulation can then be interpreted as a hypothetical scenario where additional drivers (the gig-economy delivery drivers) can potentially cause more accidents than the historical rate, while the design option choices can reduce said accidents as well as the ones present in the historical data. Two sources of public data were used. The first dataset pertains to real accident occurrences in a real city [41]. The second dataset pertains to real data for psychophysiological fatigue, as collected in [42]. The first source [41] comprises an extensive collection of historical real data for traffic accidents gathered by its authors from real-time data from MapQuest and MS-Bing [41]. It includes weather information, day of the week, language embeddings for typical traffic incidents for each region, as well as a description of the amenities available at each region. It is from this dataset that we obtain priors for accident rates and that we use for traffic simulation in general. The second source [42] is a multi-modal dataset for modeling the interplay between physical and mental fatigue and its impact on cognitive performance, where participants were subjected to physically and mentally taxing tasks while wearing smart devices. We use this dataset to simulate readings belonging to drivers at different levels of cognitive internal control.

In our simulation, drivers are separated into two cohorts: experienced drivers and new ones. New drivers comprise 20 percent of the workforce, with the remainder being

experienced drivers. Each driver has a personal maximum current level of self-regulatory control sampled from $N(\mu = 42, \sigma = 0.6)$. The difference between experienced drivers and new ones is how they deal with self-regulatory control depletion. Experienced drivers could be more susceptible to depletion due to self-regulatory fatigue because of resource loss and survival anxiety in the gig economy. In this hypothetical scenario, there could have been erratic gig-economy employment and low income, which led to the experienced delivery driver having lost financial savings, i.e., loss of resources [43].

Every timestep in the simulation represents a window of fifteen minutes of real time, as is the case in [41]. Also as in [41], the city is discretized into five-by-five kilometer regions. Deliveries are assigned to regions proportionally to the number of amenities present in said region, as provided by open street maps. Drivers will then move to their destinations using real routes, from a pre-computed set also provided by open street maps. The expected delivery time for the driver is the one provided in the route, but the actual time taken for the driver to perform each step of the route has a random component added to it to simulate variations in driving time. For experienced drivers, a number of seconds is sampled from $N(\mu = 0.0, \sigma = 1.4)$, while for the other ones, it is sampled from $N(\mu = -0.2, \sigma = 1.4)$. Each step in the route also depletes the driver stochastically, reducing its internal control by $N(\mu = 0.15, \sigma = 0.15)$ in the case of new drivers and $N(\mu = 0.2, \sigma = 0.15)$ for experienced ones. In case there are monetary penalties for each minute of delay, those values are multiplied by 1.3 to represent the more stressful environment.

If a driver is depleted (current level of internal control below 20.0) and expects to be late for the current delivery, there is a chance of breaking traffic regulations by speeding. We consider drivers as rational actors who will not speed in the presence of speed cameras (as long as those are visible to them). The extra speed bumps can also reduce the likelihood of speeding, with a two percent chance of deterring the driver (the likelihood that the extra speed bump is at the driver's location, as we place the speed bumps in five-by-five regions and not specific coordinates). Here, two percent is a heuristic value selected to represent the small area of speed bumps in relation to the five-by-five area of the grids. If those conditions are met, the driver will speed with a probability inversely proportional to its current internal control level, calculated as 100 minus c divided by 100, where c is the current internal control level for the driver. Speeding not only makes accidents more dangerous but also increases the number of accidents [44]. For calculating the number of accidents, we use the information present in [44]:

$$LO_2 = LO_1 \left(\frac{v_2}{v_1} \right)^2 \quad (1)$$

where LO_1 is the historical crash rate for that location (according to the data collected by [41], v_2 is the new average driver speed, and v_1 is the historical one, based on the Google Maps route). This allows us to calculate LO_2 , the update rate of crashes for the gig-economy drivers. For simplicity, we assume drivers can reach a maximum speed of double the original one, leading to a four times increase in the crash rate. While we rely on the same source and data interval as [41], we increase the negative sampling (presence of moments without accidents) from their original two percent of real events to twenty percent. This was performed not only to have more data for the simulation, but also to better reflect the real rate of accidents. In other words, for the "no-accident" class, we take ten times more data when compared to the original paper [41], since we focus on a more realistic simulation instead of on reducing the class unbalance that was performed for machine-learning purposes in [41]. Again, for simplicity, we do not remove drivers from the road if they have caused an accident during a specific day. This is because our simulation has a limited number of drivers.

In the implemented model, choices for Question 2 to Question 7 provide static aspects of the simulation. By contrast, choices for Question 1 and Question 8 provide dynamic inputs to the implemented world model that influence actions that happen to agents in every step of the simulation. If reminders are to be sent to drivers (Question 1), this happens every 24 timesteps and changes the current internal level of control of drivers by a number sampled from $N(\mu = 1.0, \sigma = 0.5)$. Completing deliveries increases the current internal level of control. If the delivery was achieved on time, the drivers' current internal level of control goes back to their personal maximum m , while in the case of a delay, it is set as $m((0.5 + U(0, 1))/2)$, with U being the uniform distribution. Of course, accidents are not only due to speeding. In order to include other accidents, those that go beyond the ones potentially caused by the gig-workers, we replay the real data, accounting for the new environment. The biggest factor here is the presence of speed cameras, which are known to reduce the number of traffic accidents [45]. For reducing the number of accidents, we implemented a conservative version of the statistics in [45], making a camera presence have a twenty percent chance of stopping a historical accident.

Question 8 proposes using machine learning to optimize the placement of one camera. This can be performed based on the region most likely to have an accident in the next 15 min. This means $\arg \max_r p(a|r)$, where r represents a discretized city region, and $p(a|r)$ is the probability an accident will occur at region r given its current features (weather, etc.). Alternatively, the optimization of the camera location can be performed based on the region with the most fatigued drivers, that is,

$$\operatorname{argmax}_r \sum_{d \in r} \left(\begin{cases} 1, & \text{for } p(f|s_d) > 0.5 \\ 0, & \text{else} \end{cases} \right)$$

where r is a city region, d is a driver, and $p(f|s_d)$ is the probability a driver is fatigued given a set of physiological readings. This relies on data from [42]. In that dataset, we use the information provided by empathica e4 and the labels provided by the dataset. Labels are mapped to our scenario as follows: readings for sessions with the mental fatigue label are considered possible instances of having a level of current internal control below 20, whereas all other sessions such as those labeled as baseline for fatigue are considered as data that would be sampled from drivers that have a level of current internal control above 20.

Using readings from fatigueset [42], we create instances with mapped labels and 20 input features. Instances are created using 60 s sliding windows with a 1-step size. The 20 features were extracted inside the windows using NeuroKit2 [46] for calculating variables. For temperature, we used its standard deviation, slope, minimum, maximum, and median value. In the case of heart-related features, we used mean heart rate, RMSSD, SDNN, SD2, and VHF. Regarding Electrodermal Activity (EDA), we decomposed it into tonic and phasic components. For each, we computed its standard deviation, slope, minimum, maximum, and median values as features. Each driver produces one sample of data per simulation iteration, sampled from the test set depending on its current level of internal control.

Regarding the model, for each class, an AML atomization is learned by training for 500 generations. For each class, we select all atoms with more than one constant to generate the input representation. Each atom becomes a binary feature with value 1 when the atom is present, for example, and 0 otherwise. Using its new representation, we train for each class a logistic regression classifier in order to weigh each atom's importance. For computing predictions, we select classes based on $\arg \max_c p_c(c|ar_c)$, where c is a class, p_c its learned

logistic regression classifier, and ar_c its AML representation. For computing probabilities for each class, we combine models by normalizing probabilities using

$$p(c_i | ar) = \frac{p_{c_i}(p_{c_1} | ar_{c_i})}{\sum_{c_j} p_{c_j}(c_j | ar_{c_j})}$$

where ar represents the set of atomizations for each class.

Training data in the case of accident prediction consists of the first 50 events for all regions. Regarding data, training data is all data of subjects 5, 9, and 6, while test data is sampled from the other nine subjects.

Retraining of the implemented world model, which is summarized in Figure 5, is based on the newly observed data, which is then used for training in accordance with [15]. In the case of accident prediction, this means directly training on observations of accidents so far. This includes expanding the training set and modifying the real labels if they changed due to the simulation, as in the case of a historical accident being avoided by the presence of a mobile camera.

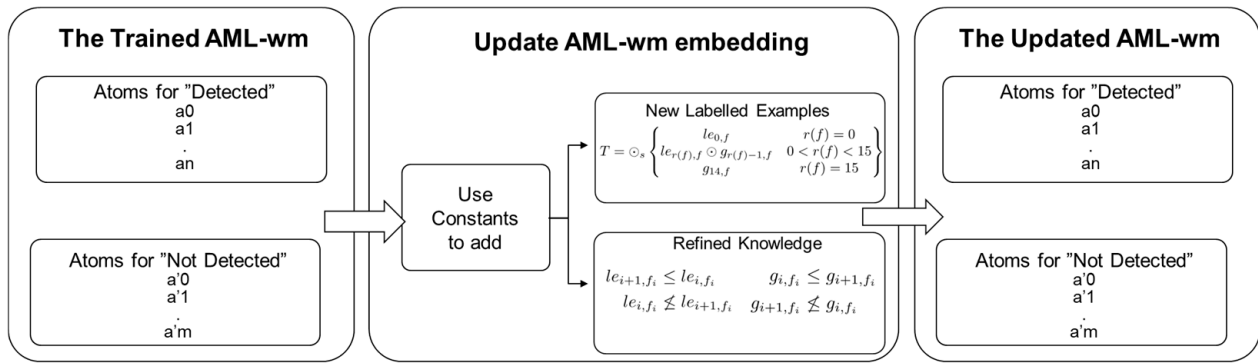


Figure 5. Updating the implemented world model for the prediction of delivery driver behavior.

For the physiological data prediction of mental fatigue, we cannot include all seen data as training input, as this would contaminate the test set (by providing the training data of test subjects). Therefore, we simply expand the training set to also include data from subjects 7, 8, and 11 and keep the other six test subjects as the test set.

5. Results

First, in this section, results are shown for when the first option is chosen for all questions. Then, results are shown for when the second option is chosen for all questions. For each question, the first option and the second option are opposites. For example, the first option for the fifth question is that delivery drivers do not have deductions from their payments if they deliver late. By contrast, the second option for the fifth question is that delivery drivers do have deductions from their payments if they deliver late. Such opposing preferences are consistent with polarization of opinions in many societies [47,48]. Subsequently, results are shown for when the implemented world model is updated to improve the prediction of where accidents are most likely to happen. This is achieved by retraining the implemented world model using the observed data in the simulation. This updating of the implemented world model is the path of least action to inform road traffic system design to reduce the probability of accidents. In particular, improving design by locating the speed camera placement where accidents are most likely to happen in order to reduce the probability of accidents actually happening.

Those first options for the eight questions are as follows: (1) all delivery drivers sent automated messages to remind them about the need to prioritize safe driving; (2) speed

bumps constructed close to a kindergarten in the city; (3) speed camera set-up close to a kindergarten in the city; (4) delivery drivers allocated the number of minutes indicated by a publicly available journey planner app; (5) delivery drivers do not have deductions from their payments if they deliver late; (6) hand-held speed cameras used where they can be seen from a far distance by drivers; (7) hand-held cameras always at the same locations; (8) access to machine learning predictions of where road traffic regulations violations are most likely to happen.

Figure 6 shows an example of real-time visualization of consequences when the first option is chosen for all eight questions. From left to right, the visualization shows the locations of accidents on a grid map, the number of accidents in relation to the number of simulation iterations, and delivery statistics in relation to the number of simulation iterations. Figure 6 shows two map grids where accidents are predicted to be most likely to occur. Also, Figure 6 shows that 742 accidents are predicted to occur based on 256 simulation iterations, while the number of on-time deliveries per iteration varies between 11 and 25.

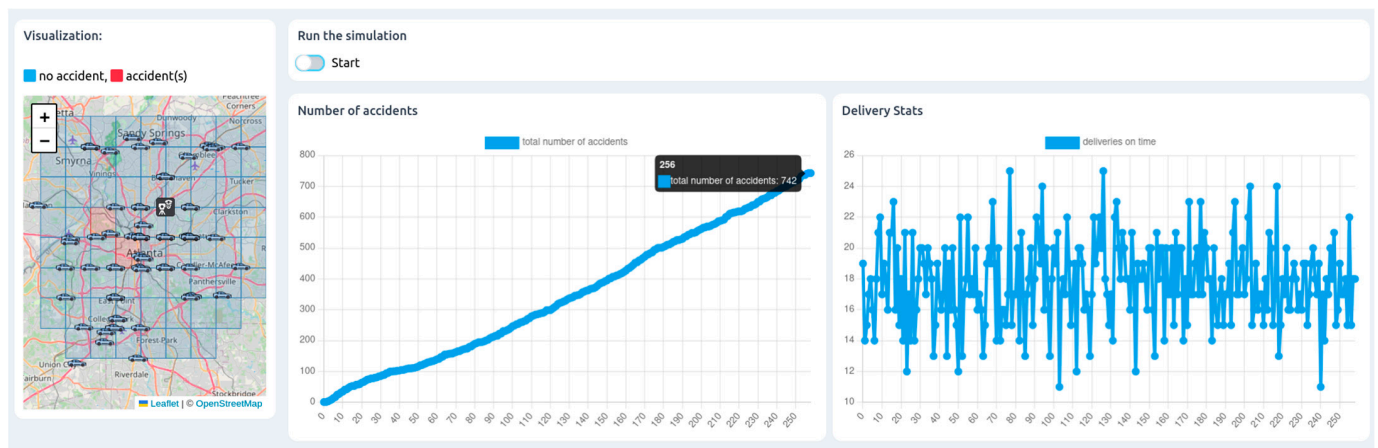


Figure 6. Simulation results after the first option for all questions has been chosen.

The second options for the eight questions are as follows: (1) reminder messages not sent to delivery drivers; (2) speed bumps constructed close to an elderly persons' home in the countryside; (3) speed camera set-up close to an elderly persons' home in the countryside; (4) delivery times as indicated by a publicly available journey planner app minus 10 percent; (5) delivery drivers have two euros deducted per minute for late deliveries; (6) hand-held speed cameras where they cannot be seen from a far distance by drivers; (7) hand-held cameras deployed at locations that often change; (8) access to machine learning predictions of which delivery drivers are most likely to violate road traffic regulations. Figure 7 shows an example of real-time visualization of consequences when the second option is chosen for all eight questions.

Figure 7 shows the two map grids where accidents are predicted to be most likely to occur. These are different from the two map grids shown in Figure 6. Also, Figure 7 shows that 1069 accidents are predicted to occur based on 256 simulation iterations. Compared to the 742 accidents predicted when the first option is chosen for all of the eight questions, this is an increase of 44 percent in the number of accidents. Also, Figure 7 shows that the number of on-time deliveries varies between zero and two. This is a notable difference from the number of on-time deliveries varying between 11 and 25 shown in Figure 6. Overall, Figure 7 summarizes that the second set of options leads not only to more accidents but also to an increase in late deliveries. This is mainly due to the choice for the fourth question to be to reduce delivery durations.

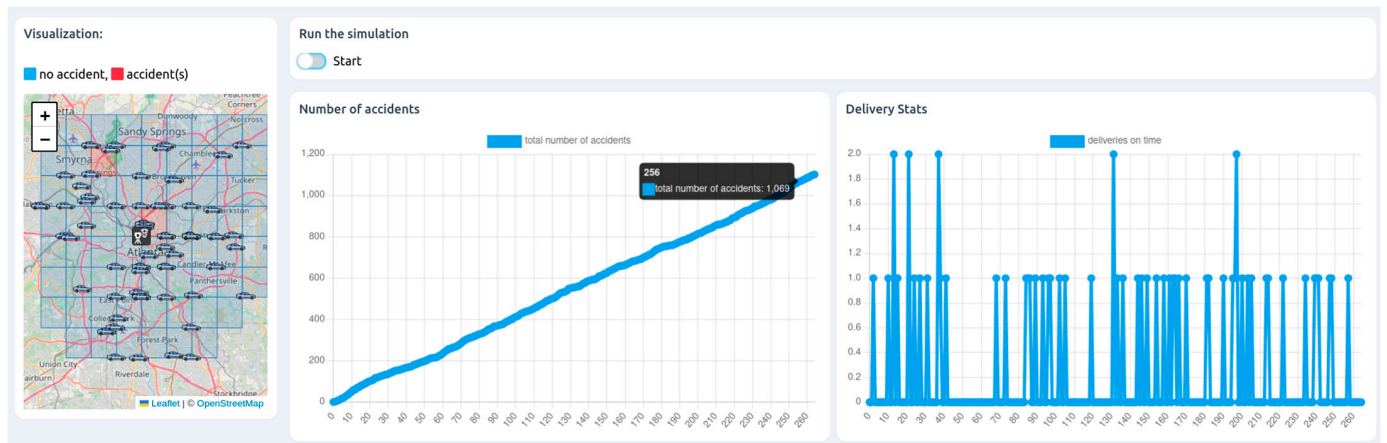


Figure 7. Simulation results after the second option for all questions has been chosen.

As summarized in Figure 5 above in Section 4, the implemented world model is retrained based on the newly collected data from the latest simulation that has been influenced by the prior choices. Figure 8 shows results after updating the implemented world model.

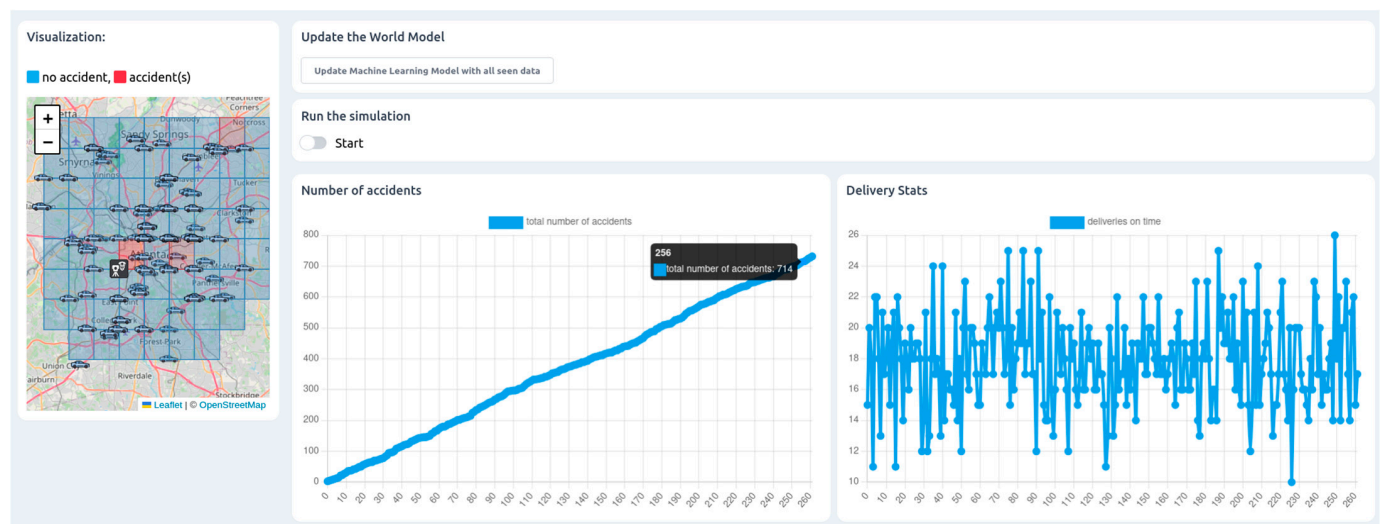


Figure 8. Simulation results after implementing the world model retrained with observed results.

Figure 8 shows that the spread of on-time deliveries increases to varying between 10 and 26. Importantly, the number of accidents decreases from 742 to 714 after the same number of simulation iterations, 256. Mainly, this is due to improvement in the implemented world model's capability to optimize the speed camera's location to predict better where accidents are likely to happen. Indicative of this improvement is three map grids, rather than two map grids in Figure 6 and in Figure 7, being shown in Figure 8 as where accidents are predicted to be most likely to occur.

Table 1 provides a comparative performance summary for the case of directly predicting if accidents will happen in a city region.

Table 1. Performance of different models for the accident prediction task based on features.

Approach	Accuracy	Macro F1
Algebraic Machine Learning with Logistic Regression	97%	53%
Logistic Regression	94%	52%

AML-based models presented better accuracy. This is due to AML models having a better precision, at the cost of lower recall in accident prediction. Since there are more cases where accidents are not happening in the test data (approximately 37 times more), this leads to higher accuracy for AML models, but overall comparable performance in terms of Macro F1.

6. Discussion

A world model has been formulated and implemented, which, because of its thriftiness, is notably different from large photo-realistic world models. The principal contribution of the reported study is to show how a thrifty world model can be sufficient to provide predictions that can inform the design of complex biosocial–technical systems. Results indicate that the thrifty world model summarized in Figure 3 is sufficient to encompass biosocial–technical complexity in predictions of where and when it is most likely that there will be accidents. The reported study has the limitation of addressing only one example: the design of road traffic systems. Nonetheless, study findings have implications for explainable AI, for artificial world models, and for frugal AI.

The explainability of machine learning models can provide a starting point for shared interpretability among people who initially can have different, even opposing, human world models and associated world views [17]. The thrifty structure for the world model summarized in Figure 3 facilitates explainability because it comprises only three categories, including a total of only eight system design choice options. Furthermore, application of the thrifty structure with machine learning, in this example with AML, facilitates shared interpretability of consequences from choices within the scope of the world model by showing the consequences arising from different choices. In particular, one person may choose all the first options for the eight choices, and see the consequences of their choices (Figure 6). By contrast, another person may choose all the second options for the eight choices, and see the consequences of their choices (Figure 7). In other words, as is often the case [47,48], different people can begin with opposing choices, such as there should not be deductions from drivers' payments if deliveries are late versus there should be deductions from drivers' payments if deliveries are late. Subsequently, both can see what choices are necessary to reduce the potential for accidents (Figure 8), irrespective of their initial choices based on their own internal human world models.

In 2025, large machine learning world models are being introduced [7–9], which can include photo-realistic environments. However, photo-realistic environments are not necessarily consistent with natural world models that are based on sensory ecologies [49], which can enable maximum information gain while minimizing information acquisition costs such as energy expenditure [50]. Here, a thrifty world model has been presented that is very far from the large photo-realistic environments being introduced by others. However, the thrifty world model presented here is consistent with natural world models in its minimalism. Thriftiness is a very valuable characteristic in the unsustainable context of soaring energy expenditure and water consumption arising from the use of AI [51–53] and the increasing recognition of the urgent need for AI to be much more frugal [18,54,55]. The term frugal refers to doing more with less, for example, for more people [56]. For machine learning, this can involve pruning models [54], increasing the generalizability of models [55], and/or focusing on efficiency in the development of systems involving machine learning [18]. The thrifty world model presented here already has a minimal structure, which does not need to be pruned. Furthermore, its three categories of internal control, external pressure, and physical environment are relevant to a wide range of complex biosocial–technical systems. Moreover, its application with machine learning

illustrates that inclusion of as few as eight choice options can efficiently encompass a wide range of alternative scenarios in the operation of a complex biosocial–technical system.

Accordingly, future research could consider thrifty world models as an alternative to large photo-realistic world models. Future research could be focused on the formulation and implementation of thrifty world models to address design problems to which large, computationally intensive world models are being applied. An interesting question for future research is to what extent the three-category model summarized in Figure 1 can provide the basis for thrifty world models for biosocial–technical system design in many different fields. The implementation of this thrifty world model with machine learning will depend on the availability of datasets. For example, datasets for human fatigue in different environments can be related to levels of internal control in relation to levels of external pressure. In other fields where environments change more slowly, longer-term simulations than those reported here will be needed.

Author Contributions: Conceptualization, investigation, writing—original draft, writing—review and editing, S.F.; software, formal analysis, writing—original draft, writing—review and editing, V.F.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Union (EU) Horizon 2020 project ALMA grant number 952091.

Data Availability Statement: The data used in this study are from [41,42].

Acknowledgments: The authors thank Fernando Martin-Maroto for technical advice.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lewin, K. Psychoanalysis and topological psychology. *Bull. Menn. Clin.* **1937**, *1*, 202–212.
2. Parkes, C.M. Psycho-social transitions: A field for study. *Soc. Sci. Med.* **1971**, *5*, 101–115. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Conant, R.C.; Ashby, W.R. Every good regulator of a system must be a model of that system. *Int. J. Syst. Sci.* **1970**, *1*, 89–97. [\[CrossRef\]](#)
4. Forrester, J.W. Counterintuitive behavior of social systems. *Technol. Rev.* **1971**, *2*, 109–140.
5. Metzinger, T. Empirical perspectives from the self-model theory of subjectivity: A brief summary with examples. *Prog. Brain Res.* **2008**, *168*, 215–245.
6. Friston, K.; Moran, R.J.; Nagai, Y.; Taniguchi, T.; Gomi, H.; Tenenbaum, J. World model learning and inference. *Neural Netw.* **2021**, *144*, 573–590. [\[CrossRef\]](#)
7. Hafner, D.; Pasukonis, J.; Ba, J.; Lillicrap, J. Mastering diverse control tasks through world models. *Nature* **2025**, *640*, 647–653. [\[CrossRef\]](#)
8. Zhu, Z.; Wang, X.; Zhao, W.; Min, C.; Deng, N.; Dou, M.; Wang, Y.; Shi, B.; Wang, K.; Zhang, C.; et al. Is sora a world simulator? A comprehensive survey on general world models and beyond. *arXiv* **2024**, arXiv:2405.03520. [\[CrossRef\]](#)
9. Agarwal, N.; Ali, A.; Bala, M.; Balaji, Y.; Barker, E.; Cai, T.; Chattopadhyay, P.; Chen, Y.; Cui, Y.; Ding, Y.; et al. Cosmos world foundation model platform for physical AI. *arXiv* **2025**, arXiv:2501.03575. [\[CrossRef\]](#)
10. de Maupertuis, P.L.M. Les lois du mouvement et du repos déduites d’un principe métaphysique. In *Histoire de l’Académie Royale des Sciences et des Belles—Lettres de Berlin*; Haude: Berlin, Germany, 1746; p. 286.
11. Eckmann, J.P.; Tlusty, T. Dimensional reduction in complex living systems: Where, why, and how. *BioEssays* **2021**, *43*, 2100062. [\[CrossRef\]](#)
12. Kaila, V.R.; Annala, A. Natural selection for least action. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2008**, *464*, 3055–3070. [\[CrossRef\]](#)
13. Feynman, R.P. The Principle of Least Action in Quantum Mechanics. Ph.D. Thesis, Princeton University, Princeton, NJ, USA, 1942.
14. Coopersmith, J. *The Lazy Universe: An Introduction to the Principle of Least Action*; Oxford University Press: Oxford, UK, 2017.
15. Martin-Maroto, F.; de Polavieja, G.G. Algebraic machine learning. *arXiv* **2018**, arXiv:1803.05252. [\[CrossRef\]](#)
16. Haidar, I.M.; Sliman, L.; Damaj, I.W.; Haidar, A.M. Legacy Versus Algebraic Machine Learning: A Comparative Study. In *International Congress of Electrical and Computer Engineering*, 2nd ed.; Seyman, M.N., Ed.; ICECENG 2023; EAI/Springer Innovations in Communication and Computing; Springer: Cham, Switzerland, 2024.
17. Fox, S.; Rey, V.F. A Cognitive Load Theory (CLT) Analysis of Machine Learning Explainability, Transparency, Interpretability, and Shared Interpretability. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 1494–1509. [\[CrossRef\]](#)

18. Peixoto, E.; Torres, D.; Carneiro, D.; Silva, B.; Novais, P. Efficient MLOps: Meta-learning meets Frugal AI. In *Advances in Artificial Intelligence in Manufacturing II*; Alexopoulos, K., Makris, S., Stavropoulos, P., Eds.; ESAIM 2024; Lecture Notes in Mechanical Engineering; Springer: Cham, Switzerland, 2024. [\[CrossRef\]](#)
19. Fox, S.; Griffy-Brown, C.; Dabic, M. From socio-technical systems to biosocial technical systems: New themes and new guidance for the field of technology in society. *Technol. Soc.* **2020**, *62*, 101291. [\[CrossRef\]](#)
20. de Campos, A.S.; Feltrin, R.B.; Pamplona da Costa, J.O.; Dos Santos, P.X.; Carvalho, G.; de Medeiros, M.B.; Chagnon, P. Biosocial Technical Systems: An emerging approach to analyse responses to novel diseases. In *The COVID-19 Aftermath: Volume II: Lessons Learned*; Rezaei, N., Ed.; Springer Nature: Cham, Switzerland, 2024; pp. 303–313.
21. Costa, T.G.; de Oliveira, V.N.; Santos, D.A.T.; Viana, R.B.; Andrade, M.S.; Vancini, R.L.; de Lira, C.A.B. The burden of prolonged sedentary behavior imposed by uberization. *Sports Med. Health Sci.* **2023**, *5*, 159–163. [\[CrossRef\]](#)
22. Park, J.H.; Moon, J.H.; Kim, H.J.; Kong, M.H.; Oh, Y.H. Sedentary lifestyle: Overview of updated evidence of potential health risks. *Korean J. Fam. Med.* **2020**, *41*, 365. [\[CrossRef\]](#)
23. Wang, Y.; Wang, G.; Chen, Q.; Li, L. Depletion, moral identity, and unethical behavior: Why people behave unethically after self-control exertion. *Conscious. Cogn.* **2017**, *56*, 188–198. [\[CrossRef\]](#)
24. Evans, D.R.; Boggero, I.A.; Segerstrom, S.C. The nature of self-regulatory fatigue and “ego depletion” lessons from physical fatigue. *Personal. Soc. Psychol. Rev.* **2016**, *20*, 291–310. [\[CrossRef\]](#)
25. Lee, E.J.; Yun, J.H. Moral incompetency under time constraint. *J. Bus. Res.* **2019**, *99*, 438–445. [\[CrossRef\]](#)
26. Umphress, E.E.; Bingham, J.B.; Mitchell, M.S. Unethical behavior in the name of the company: The moderating effect of organizational identification and positive reciprocity beliefs on unethical pro-organizational behavior. *J. Appl. Psychol.* **2010**, *95*, 769. [\[CrossRef\]](#)
27. Christie, N.; Ward, H. The health and safety risks for people who drive for work in the gig economy. *J. Transp. Health* **2019**, *13*, 115–127. [\[CrossRef\]](#)
28. Popan, C. Embodied precariat and digital control in the “gig economy”: The mobile labor of food delivery workers. *J. Urban Technol.* **2021**, *31*, 109–128. [\[CrossRef\]](#)
29. Bazerman, M.H.; Gino, F. Behavioral ethics: Toward a deeper understanding of moral judgment and dishonesty. *Annu. Rev. Law Soc. Sci.* **2012**, *8*, 85–104. [\[CrossRef\]](#)
30. Houdek, P. Fraud and understanding the moral mind: Need for implementation of organizational characteristics into behavioral ethics. *Sci. Eng. Ethics* **2020**, *26*, 691–707. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Fox, S. Behavioral ethics ecologies of human-artificial intelligence systems. *Behav. Sci.* **2022**, *12*, 103. [\[CrossRef\]](#)
32. Hursthouse, R.; Pettigrove, G. Virtue Ethics. In *Stanford Encyclopedia of Philosophy*; Edward, N.Z., Uri, N., Eds. Available online: <https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue> (accessed on 22 May 2025).
33. Alexander, L.; Moore, M. Deontological Ethics. In *Stanford Encyclopedia of Philosophy*; Edward, N.Z., Uri, N., Eds. Available online: <https://plato.stanford.edu/entries/ethics-deontological> (accessed on 22 May 2025).
34. Sinnott-Armstrong, W. Consequentialism. In *Stanford Encyclopedia of Philosophy*; Edward, N.Z., Uri, N., Eds. Available online: <https://plato.stanford.edu/entries/consequentialism> (accessed on 22 May 2025).
35. Baranzke, H. “Sanctity-of-Life”—A Bioethical Principle for a Right to Life? *Ethical Theory Moral Pract.* **2012**, *15*, 295–308. [\[CrossRef\]](#)
36. Shafi, S.; Gentilello, L. A nationwide speed limit ≤ 65 miles per hour will save thousands of lives. *Am. J. Surg.* **2007**, *193*, 719–722. [\[CrossRef\]](#)
37. Orr, W.; Henne, K.; Lee, A.; Harb, J.I.; Alphonso, F.C. Necrocapitalism in the gig economy: The case of platform food couriers in Australia. *Antipode* **2023**, *55*, 200–221. [\[CrossRef\]](#)
38. Antić, B.; Pešić, D.; Vujanić, M.; Lipovac, K. The influence of speed bumps heights to the decrease of the vehicle speed—Belgrade experience. *Saf. Sci.* **2013**, *57*, 303–312. [\[CrossRef\]](#)
39. Knapton, S. Speed bumps could be removed to cut traffic pollution and save lives. *The Telegraph*. 1 December 2016. Available online: <https://www.telegraph.co.uk/news/2016/12/01/speed-bumps-could-removed-cut-traffic-pollution-save-lives/> (accessed on 22 May 2025).
40. Wilson, C.; Willis, C.; Hendrikz, J.K.; Le Brocque, R.; Bellamy, N. Speed cameras for the prevention of road traffic injuries and deaths. *Cochrane Database Syst. Rev.* **2010**, *11*, CD004607.
41. Moosavi, S.; Samavatian, M.H.; Parthasarathy, S.; Teodorescu, R.; Ramnath, R. Accident risk prediction based on heterogeneous sparse data: New dataset and insights. In Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 5–8 November 2019; pp. 33–42.
42. Kalanadhabhatta, M.; Min, C.; Montanari, A.; Kawsar, F. FatigueSet: A multi-modal dataset for modeling mental fatigue and fatigability. In *Pervasive Computing Technologies for Healthcare*; Lewy, H., Barkan, R., Eds.; PH 2021; Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering; Springer: Cham, Switzerland, 2022; Volume 431. [\[CrossRef\]](#)

43. Hobfoll, S.E. Conservation of resources theory: Its implication for stress, health, and resilience. *Oxf. Handb. Stress Health Coping* **2011**, *127*, 147.
44. Nilsson, G. The effects of speed limits on traffic accidents in Sweden. In Proceedings of the International Symposium on the Effects of Speed Limits on Traffic Accidents and Transport Energy Use, Dublin, Ireland, 6–8 October 1981; OECD: Paris, France, 1981; pp. 1–8.
45. Pérez, K.; Marí-Dell’Olmo, M.; Tobias, A.; Borrell, C. Reducing road traffic injuries: Effectiveness of speed cameras in an urban setting. *Am. J. Public Health* **2007**, *97*, 1632–1637. [[CrossRef](#)]
46. Makowski, D.; Pham, T.; Lau, Z.J.; Brammer, J.C.; Lespinasse, F.; Pham, H.; Scholzel, C.; Chen, S.H.A. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behav. Res. Methods* **2021**, *53*, 1689–1696. [[CrossRef](#)] [[PubMed](#)]
47. Cely, T. New threat of culture wars? The religious roots of public opinion polarization on morality issues in Europe. *Eur. Union Politics* **2025**, *26*, 226–254. [[CrossRef](#)]
48. Revers, M. Performative polarization: The interactional and cultural drivers of political antagonism. *Cult. Sociol.* **2025**, *19*, 207–228. [[CrossRef](#)]
49. Stevens, M. *Sensory Ecology, Behaviour, and Evolution*; Oxford University Press: Oxford, UK, 2013.
50. Schneeberger, K.; Taborsky, M. The role of sensory ecology and cognition in social decisions: Costs of acquiring information matter. *Funct. Ecol.* **2020**, *34*, 302–309. [[CrossRef](#)]
51. Crawford, K. Generative AI’s environmental costs are soaring—And mostly secret. *Nature* **2024**, *626*, 693. [[CrossRef](#)]
52. Berthelot, A.; Caron, E.; Jay, M.; Lefèvre, L. Estimating the environmental impact of Generative-AI services using an LCA-based methodology. *Procedia CIRP* **2024**, *122*, 707–712. [[CrossRef](#)]
53. Zewe, A. Explained: Generative AI’s environmental impact. *MIT News*. 17 January 2025. Available online: <https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117> (accessed on 27 May 2025).
54. Klimczak, A.; Wenka, M.; Ganzha, M.; Paprzycki, M.; Mańdziuk, J. Towards frugal artificial intelligence: Exploring neural network pruning and binarization. In *International Symposium on Intelligent Informatics*; Thampi, S.M., Mukhopadhyay, J., Paprzycki, M., Li, K.C., Eds.; ISI 2022; Smart Innovation, Systems and Technologies; Springer: Singapore, 2023; Volume 333. [[CrossRef](#)]
55. Martínez-Arellano, G.; Ratchev, S. Towards Frugal Industrial AI: A framework for the development of scalable and robust machine learning models in the shop floor. *Int. J. Adv. Manuf. Technol.* **2025**, *138*, 169–191. [[CrossRef](#)]
56. Prabhu, J. Frugal innovation: Doing more with less for more. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2017**, *375*, 20160372. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.