# BeliefNet: A neurosymbolic model for context-based traversability predictions in complex environments

Tom Scott, Argyrios Zolotas and Yang Xing
*Centre for Autonomous and Cyber-Physical Systems, Cranfield University, UK*

**Abstract.** Knowing how to traverse complex unstructured environments is a difficult and multivariate challenge, but one which humans can achieve through logic, reasoning, and experience; yet some of the most beneficial use cases for autonomous systems require them to operate effectively in complex environments without regular or significant human intervention. Furthermore, for machines to support humans in some of the more critical use cases, trust in decision making will be crucial, ensuring operators have confidence to deploy the capabilities. Despite its importance, enabling autonomous agents to navigate effectively and reliably in complex terrain is a difficult and unsolved challenge. Advances in neurosymbolic AI present an opportunity to significantly enhance performance in complex, explainable, and uncertain decision making, such as autonomous traversability analysis, by drawing together symbolic reasoning with the learning capability of neural networks. The challenge of complex environments is complicated by its non-deterministic nature; terrain will adapt and change through domains, and its properties can adapt rapidly based on external factors like weather or objects that are in proximity, which is true for one location on one day, will not persist. This paper presents a new neurosymbolic model structure that was designed specifically for this task. It uses experience to build a world model, similar to that of a neural network, but with some key delineating features such as full explainability, through life adaptation or evolution, and zero-shot capability, enabling it to perform as both a reasoning engine and a memory representation for an autonomous system. This provides the reasoning backbone for an autonomous agent to determine the level of risk each object presents based on its context and therefore determine the best possible route.

Keywords: Neurosymbolic AI, Machine learning, Knowledge-based learning, Autonomous systems, Complex environments

## 1. Introduction

Autonomous systems present an opportunity to transform the way humans complete some of the most dangerous, unpleasant, or persistent tasks, especially within domains such as Defence or Search and Rescue. These use cases present some of the greatest beneficiaries of autonomous systems, but have some of the most demanding requirements, most notably the ability to operate reliably in very complex terrain and dynamic domains, whilst maintaining a high degree of trust by their operators to complete their task at hand. Robustly operating in complex environments requires platforms to operate in both unstructured and uncertain terrain, where clear transition points between features may not exist, with high variation in slope, roughness and unpredictable terrain features like holes or depressions [67], [70], [71]. Furthermore, the characteristics of an object cannot be determined effectively without understanding the context in which it is found. Navigating requires inductive and deductive reasoning, an understanding of the environmental conditions, probabilistic judgement, and the ability to handle uncertainty. When considering autonomous agents, neither a symbolic nor a neural approach replicates them all sufficiently. Neural approaches generally fail to reason effectively and suffer from a lack of explainability but can be adaptive to out-of-distribution data, whilst symbolic approaches can reason but require a significant upfront knowledge base and

cannot effectively generalise. Fundamentally, performing these activities within an autonomous platform is not a simple extrapolation of either approach. The traversal of complex environments remains an outstanding challenge in the field of autonomous systems [17], [74].

Navigating complex terrain can be considered across a number of fields of research, such as perception, localisation, cognition and motion control [57]. This paper focuses on cognition and specifically on how to enable an agent to determine the traversability of a target object. When considering the prediction of traversability for a given object within complex environments, the continuous, layered structure of individual objects means that the assessment of an object in isolation is insufficient for making an accurate prediction. The images in Figure 1 show an example of two separate examples of the same trail object, both of which have separate performances, caused as a result of the adjacent objects. Furthermore, terrain characteristics are not consistent in all domains and environments; the performance of grass changes if it has rained in the last hour, and in winter, this may be the last day. As a result, such environmental information is the context and parameters of the overall prediction. Consideration must also be made to the rapid domain evolutions and inconsistencies required when operating within such an environment, meaning traversability predictions must have a broad generalisation capacity, enabling routine handling of previously unseen situations.

Vehicles will operate in close proximity to humans who, for both safety and functional reasons, need confidence in actions and to understand why a decision is to be made. As a result, operator trust must be considered within any agent cognition. This constraint makes any potential for context prediction within a conventional neural network-based solution [30], [77], [20], [76] a barrier to practical deployment, as any decision justification would be concealed within the black-box nature of the model, inhibiting explainability. Conversely, explicit reasoning and logic are easier to interpret than neural architectures [78], as reasoning transcends the originator; it is communicable and can be understood externally [52]. Through explicit reasoning, operators can interpret decisions and understand errors, making the prediction more deterministic and increasing trust [26].

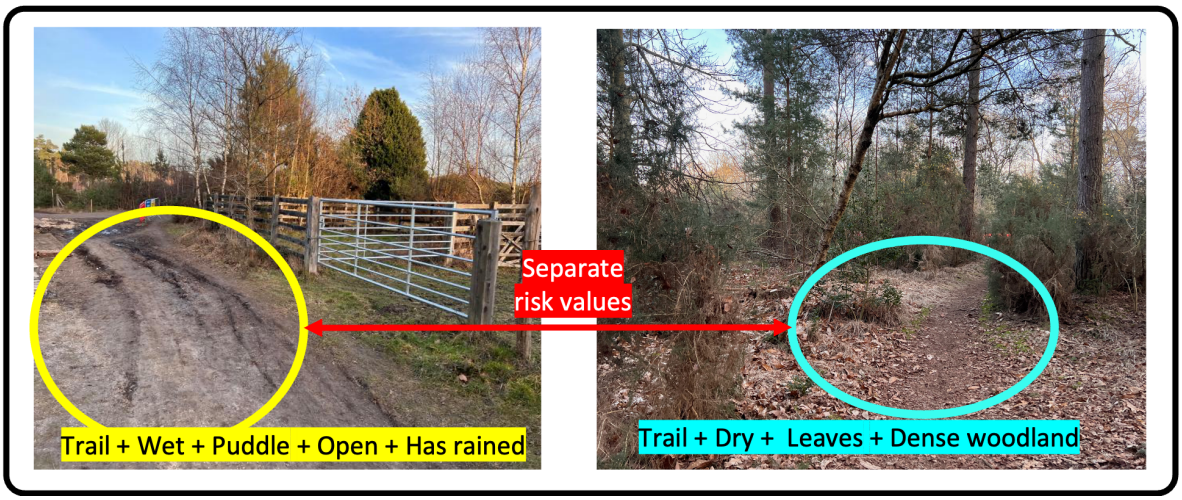


Fig. 1. An example of two instances of trail objects with their surrounding context, resulting in separate risk assessments

This concept paper builds on the concept of the World Model [42], using neurosymbolic methods [34] to develop a human-like approach to solving the challenge of autonomy in complex environments. It presents a model structure which enables an agent to make traversability predictions that account for an object's context, learn dynamically with new experiences, and use causal relationships to generalise across evolving domains. This paper outlines BeliefNet, a model designed to support explainable context-based prediction for complex environments.

BeliefNet uses a symbolically built neural architecture to form experience-based beliefs, overtime generating causal relationships between a target object, its context and traversability risk. The model is designed to train through life, learning with an agent's experience, enabling adaptation to new environments. The model seeks to extrapolate

causal relationships, enabling it to generalise effectively to different domains. It uses belief-based inference to form deterministic and explainable predictions, improving prediction accuracy, domain adaptation, and enhancing operator trust.

The contributions of this paper are as follows.

- The proposal of BeliefNet, a Neuro[Symbolic] model structure for context-based traversability prediction for autonomous systems.
- A demonstration of the performance of BeliefNet in an adapted version of the Yamaha CMU [79] dataset, to increase the performance of agent cognition in complex environments.
- Comparison of context-based terrain traversability prediction and object-based prediction.
- A high-level traversability taxonomy for ground platforms based on risk and speed.

## 2. Existing work

### 2.1. Traversability assessment

The field of traversability has received significant attention in recent years, leading to the development of three primary approaches emerging to conduct traversability assessment: terrain mapping, terrain classification, and end-to-end solutions [6]. Lidar analysis has been used extensively in traversability mapping approaches, both in direct obstacle avoidance [72][40], or in more complex feature segmentation [24] [2] [80]. Whilst delivering promising results, such approaches are spatial in nature, potentially over simplifying the traversability calculus by ignoring the environmental and situational semantics. Furthermore, the active nature of Lidar presents challenges in use cases where light emissions have negative secondary effects.

Terrain classification presents a method to incorporate semantics. Advances in computer vision, with the introduction of models such as YOLO [63] and approaches such as vision transformers [16] and panoptic segmentation [81], have made this increasingly feasible, allowing real-time inference on edge-based devices. The use of computer vision enables terrains to be segmented into constituent objects, from which semantic labels and classes can be subsequently assigned. As the terrain classification of complex environments is non-trivial, resulting from the discontinuous nature of objects, feature overlap, and environmental conditions [49], this continues to be an area of active research [12], [18], [25], [11], [75] [2], [71], [19]. Vision and Lidar modalities have been combined to integrate visual semantics with the spatial representations of Lidar [51].

Terrain classification is formed of two distinct components, first detecting and isolating a specific object within the scene, then assessing the traversability of the object. Although some, such as [2] have integrated both components, most of the research focuses mainly on accurately determining the object, not assessing the traversability. One challenge in this approach is that it can neglect the need to consider the environment and context of a specific object, ignoring that some objects will directly impact the traversability of others. Without such context, it can be challenging to make an accurate and reasoned determination, which is exacerbated as the complexity of the environment increases.

End-to-end deep learning approaches have had success in classifying the traversability of an image [77], [43]. Self-supervised approaches, in which a platform trains a model based on self-extracted features to predict the traversability of the terrain [66] [68], reduce the volume requirements for labelled data to some degree. Such approaches can be limited in generalisation performance and crucially limit explainability due to the conventional neural architecture of end-to-end deep learning approaches. The use of image segmentation, coupled with self-supervised learning, presents a method to increase explainability, but the computational requirements are prohibitive [30] and the explicit impact of object context is not explicit and uncertain. Although research into traversability assessment for complex environments has been significant, it remains an open area of research and one in which significant advances are required to enable autonomous systems to complete the desired tasks.

## 2.2. Agent cognition

Agent cognition within the field of robotics to model actions with high levels of uncertainty has had significant success using probabilistic methods. Markov Logic Networks (MLN) in event modelling have been used successfully to classify events from images based on their context [33], but suffer in complex environments due to cross-domain adaptation, handling partial rule activation, and probabilistic complexity [13], [35]. Markov Decision Processes (MDP) have also been used in path planning and path trajectory, with Partially Observable MDPs (POMDP) used to handle increasing levels of uncertainty and complex environments [41]. Belief states have been integrated into POMDP to extend planning horizons and mitigate the impact of partial observability [31], [21]. Applications in robotics and autonomous systems focus primarily on action selection. In which state estimation is used to determine the best action to take across a given planning horizon, formulation of the reward for a given action (such as its predicted traversability value) is often adjacent to the model. Probabilistic programming using tools such as ProbLog has sought to overcome the complexity limitations faced by MDPs and MLNs, such as probability modelling and handling uncertainties in predictions [56], [73], but requires a firm logical foundation, which may not be easily available in complex environments.

The concept of a belief within Bayesian epistemology considers that beliefs are not consistent, and the degree to which a belief is believed is adapted to the available body of evidence [45]. Pearl emphasises the centrality of causality in beliefs, expanding beyond correlation [58]. Pearl introduces a three-layer model that supports causality in the concept of machine learning: association, intervention, and counterfactual; progression through these layers supports the classification of causal information and improves the degree of confidence in a given belief [59]. Graph structures have presented a method to effectively communicate causal relationships between entities, noting that such relationships are not hierarchical or linear [46]. The application of causality to machine learning has recently been identified as a method to significantly increase generalisation and cross-task adaptation. It presents a complex challenge due to the nature of feature extraction from data. Scholkopf et al. presented a number of potential approaches, such as self-supervised learning and reinforcement learning. This research outlined the importance of observation and intervention in learning causal relationships [65]. The applications of causality have been applied to learn causal relationships and apply them within inference [82], [62], [84]. Causality within AI presents significant promise, though limited within this challenge by the data volumes required by approaches such as autoencoders. The application of a priori knowledge to autoencoders has been applied by Komanduri et al. as a method of reducing upfront data requirements [38].

Advances in probabilistic modelling and decision making have had a significant impact on robotics, but the high levels of variability in object class, prediction confidence, object separation, and adaptive domains seen in complex environments make logical grounding and the application of finite rule sets insufficient to effectively model the complexity. The consideration of belief states and partial observability is very relevant to context-aware traversability. The state estimation is considered not as the state of the agent given an action, but estimation of the state of beliefs held about an object, it's context, and the resulting traversability. The ability to make predictions with partially observed inputs and an incomplete understanding of object interactions, whilst being able to update understanding when new information is available, is very pertinent to complex environments and domain adaptation. The concepts of beliefs and causality are also of benefit to this problem set, supporting an agent learning from new experiences, and supporting generalisation when facing uncertainty, both of which are common in complex environments.

## 2.3. Neurosymbolic AI

Neurosymbolic AI is a promising area of research in machine decision making, explainability, and reasoning [8]. This area of research presents architectures to integrate the reasoning performance of symbolic reasoning and the learning power of sub-symbolic, connectionist or neural network-based approaches [34], based on the system-1 / system-2 approach defined by Kahneman [32]. The field is still growing and there remains diversity in approaches, but all have in common the structure of perception, integrated with existing knowledge [69], and their explainability and reasoning performance make them particularly beneficial in use cases with high levels of human-machine interaction [4]. Within the broader field of autonomous system navigation, neurosymbolic architectures have been used

to integrate physics rules into a neural network to determine the vehicle dynamics required to traverse a given path [85]. Neurosymbolic architectures are commonly represented by six core approaches [15], [34], two of particular relevance to this article are de ned as NEURO;SYMBOLIC and NEURO[SYMBOLIC] [37].

NEURO;SYMBOLIC represents a system in which symbolic and neural systems work in concert with each other, communicating and passing information between them, to achieve a common objective [34]. Examples of this are knowledge graph integration with neural networks [15] that allow a neural network to query, input to, and validate symbolic knowledge graphs. NSNnet, which passes between neural and symbolic modules in an attempt to solve hand-written Sudoku challenges, presents a unique perspective that maps both input and output to a non-symbolic output, with a central symbolic reasoning engine [1]. Both examples are dependent on a core level of symbolic reasoning. The Neuro-Symbolic Concept Learner (NCSL) is designed to unify text and visual concepts through the learning of image and question-answer pairs [50]. This model presents an interesting advance as it enables symbolic concepts to be learned, without implicit knowledge being de ned upfront.

In contrast, the NEURO[SYMBOLIC] system is one in which a neural network learns to reason about relationships between neural entities [34] [39], in effect forming a neural network of symbolic entities. This is perhaps the most complex and least mature area of research within the eld. Logic Tensor Networks (LTN) and Logic Neural Networks (LNN), which form networks from symbolic relationships and enable weighted training of the relationships using back-propagation based on a set of rst-order logic statements [3] [64]. The pLogicNet model mostly precedes the core de nitions of neurosymbolic AI and represents a method similar to Logic Tensor Networks based on the application of Markov Logic Networks [61]. The LTN and pLogicNet are designed to improve, validate, or decon ict a set of a priori logical statements. The challenge with these approaches when applied to an agent-based approach is that they require upfront knowledge that may not be practical to achieve. Models such as the Neuro Symbolic Reinforcement Learner, INSIGHT, by Luo et al. use a neural network to learn symbolic policies that support the agent in its decision-making, enabling reasoning to be learned from the environment [48].

Neurosymbolic systems have shown signi cant promise in vision and multimodal tasks, such as visual question answering and scene graph generation [44], [5], [29]. Despite success in these areas, multimodal neurosymbolic systems remain challenged in ensuring consistency between modalities as the deployment domain evolves [47]. From an autonomous system perspective, this might have the most impact at the point that multiple sensors are combined into a single collaborative neurosymbolic architecture, such as the camera and Lidar, as shown previously, a common approach. Although this remains an open challenge, it does limit the breadth of neurosymbolic architecture application in autonomous vehicles.

The current state of neurosymbolic AI presents signi cant advances in both reasoning and explainability, the NEURO[SYMBOLIC] concept of a single neural network which encapsulates symbolic reasoning presents an opportunity to represent an agent's world model. As with probabilistic approaches, current methods often rely on a set of a priori logic statements, leading to similar constraints on domain adaptation. As a result, BeliefNet has taken the concept of a symbolic network trained using sub-symbolic approaches, but in a manner that re ects the domain learning capabilities of models such as INSIGHT or the NSCL in which beliefs can be inferred from training. In the generation of beliefs, in opposition to rules, BeliefNet provides the ability to learn continuously from an agent's experience, avoiding the constraint of domain-speci c logic that fails to support out-of-distribution inference.

## 3. Approach

### 3.1. Overview

BeliefNet is a graph-based directed network in which nodes represent symbolic information, and unlike a neural network, the edges are not fully connected, but instead form relationships based on observation and counterfactual evidence. The network nodes and edges then act as neurons and connections in a neural network supporting weight optimisation. This structure enables the model to make traversability adaptations with very small amounts of data when compared with a conventional neural network, whilst retaining absolute explainability in the model's deduction. The relationship between a given set of input predicates and output results represents a belief within the model. Beliefs are something the system has some degree of con dence in being true [55], based on its own experiences.

Conceptually, human beliefs continually evolve and adapt to our experiences and our current domain, and we learn through life. When we face something unknown, we find the set of closest beliefs, use them to make a prediction, then create a new belief that captures the separation between the prediction and the truth, often captured within the concept of Predictive Coding [54]. It is this function that the BeliefNet model looks to model; conventional neural networks struggle with this approach requiring full validation after each evolution. In contrast, the graph structure of BeliefNet makes domain adaptation and counterfactual generation a deterministic function of the model throughout its use.

The BeliefNet model is designed to operate post perception, so it can be agnostic to the object classification model, or even the modality. It is also capable of integrating new predicates into the model, which means new classes can be added to a perception model, and these will be incorporated into BeliefNet as they are experienced. The model is built logically before training, in which connections between objects, their context, or existing relationships (in the case of counterfactuals) are generated dynamically. After which, a forward pass through the model is made, followed by optimisation and back-propagation using a conventional loss function. This can be achieved using a conventional upfront training set and continued further as the agent experiences its environment, providing an intervention mechanism overtime generating causal beliefs. Combined with the symbolism retained within each node, the structure provides the ability to activate only relevant sections of the model during inference, aiding explainability and providing reasoning in unknown situations. This approach acts as a zero-shot domain adaptation model, without the need for the high data volumes conventionally required through existing zero-shot approaches.

The model is designed for human interaction. The symbolic nature of the nodes and deliberate relationships means that any prediction can be traced through the model directly and that contributing nodes can be clearly identified. This enables operators to interact with the agent's cognition in novel methods, which are likely to significantly enhance trust. Operators can clearly determine why a decision was made, and can actively correct the result and use this to directly train the model. Furthermore, if they hold logic that had not yet been experienced by the model, this can be integrated as testimonial knowledge within the network directly. As a result, BeliefNet presents an approach capable of context-based traversability prediction and the ability to generate trust between the agent and the operator.

## 3.2. High level structure

The model is formed of a number of components, some of which are adaptations of existing deep-learning approaches and some which are specific to BeliefNet. At a high level, the model should be considered as post-processing of a perception model. It initialises by taking the perception output and transforming this into a graph structure, known as an instance graph. The instance graph is generated as the output of a semantic segmentation model, such as YOLO [28]. The predictions are further enhanced through a depth perception model [7], estimated 3D separation between objects, and augmented with environmental tags that represent the weather, light, and domain. The instance graph is a dense symbolic representation of a given image. During training the instance graphs are converted to a series of context graphs, representing a target object and the surrounding objects, distances, and environmental tags for a given object for which a prediction is made. Context graphs are passed to the building algorithm, which is a custom training method designed to extract causal relationships between objects, context, and a traversability value. This will occur even during inference, enabling new relationships to be formed as they are identified. This forms the basis of the network; each node has an activation function, bias parameter, and each edge has a connection weight. To make a prediction, the predicates within the context graph are activated and propagated through relationships in the graph, generating values at the output nodes. When training or when provided with feedback, the optimisation step occurs, which uses conventional back-propagation, such as the Adam algorithm [36] to adapt the weights in a supervised manner. It is the combination of the logical build process before the back-propagation which provides the reasoning capacity and explainability of the structure. The architecture in Figure 2 visually shows how these components fit together within the model.