Causal Graph Modelling with Deep Neural Engines for Strong Abstract Reasoning in Language and Vision

Gaël Gendron
University of Auckland
gael.gendron@auckland.ac.nz

Abstract
Deep learning (DL) relies on discovering correlation patterns in low-level data and aggregating the information to solve a task. Despite success in a wide variety of applications, ranging from natural language to vision tasks, the learned patterns are often brittle and do not transfer out of the training data distribution (i.e., to different domains). Causality theory proposes methods to discover and estimate cause-effect relationships beyond correlations. Its powerful inference frameworks have been recently highlighted as a potential way to improve the lack of out-of-distribution generalisation in deep neural networks. However, their applications to deep learning problems remain largely under-explored. Our work attempts to bridge this gap and apply causal graphical models to abstract and causal reasoning problems in natural language and vision, requiring strong generalisation abilities beyond correlations. We integrate causal graph modelling methods into deep vision networks and Large Language Models to improve their capacity to perform strong and out-of-distribution reasoning on complex abstract problems.

1 Introduction and Motivations
Deep learning (DL) encompasses broad areas of research based on the discovery of correlation patterns from low-level data. These models have induced breakthroughs in image and text understanding. However, one shortcoming is their prominence to overfit to a particular distribution rather than generating patterns that can generalise further [Goyal and Bengio, 2022]. Large Language Models (LLMs) have further shown impressive performance on text reasoning tasks. Still, their training on extensive data raises the question of whether they generalise or merely memorise already-seen patterns. Recent studies showed that benchmarks used to evaluate LLMs have leaked into their training, altering evaluation [Li and Flanigan, 2024]. This issue is of prominent importance in reasoning tasks, where the model must discover new unseen patterns, and abstract reasoning offers a way to evaluate it rigorously. Abstract reasoning tasks consist of recognising patterns in images, texts or symbols from only a handful of examples and extrapolating from them [Gendron et al., 2023a]. Causal models are widely used in many domains of science to extrapolate results to unseen groups or environments but have long evolved separately from DL. They allow learning cause-effect relationships and not only correlations. They are more invariant to distribution changes as they discard spurious correlations from their reasoning paths [Pearl, 2009; Schölkopf et al., 2021].

2 Contributions
We apply causal graph modelling techniques to deep learning architectures to improve their strong reasoning abilities on abstract reasoning tasks. The contributions of this project are twofold as we focus on two modalities: text and images. We study two families of architectures for their respective modalities: Large Language Models and graph networks (encompassing Graph Neural Networks (GNNs) and Variational Graph Autoencoders). The fast growth of Large Language Models in the past two years and their impressive performance on commonsense causal reasoning raised the question: do LLMs already contain the fundamental blocks for strong abstract and causal reasoning? To answer this question, we have created a benchmark for LLM evaluation on abstract reasoning tasks. We have shown that, despite their impressive performance on language modelling tasks, Large Language Models do not currently have the ability to reason abstractly
and causally [Gendron et al., 2023a]. We have investigated a solution to mitigate this problem by dividing the LLM inference process into independent causal modules learning specific mechanisms with little to no interaction, thus limiting spurious correlations in the data and improving causal and abstract reasoning [Gendron et al., 2024]. We showed that this can effectively create specialised modules with better out-of-distribution generalisation than their dense counterparts. Figure 2b describes the model. Vision-based reasoning poses different challenges than language. Images are composed of a grid of pixels with high dimensions and low-level information content. Therefore, the first step towards strong image-based reasoning is building high-level causal variables that can represent semantic concepts. We have developed a causal variational autoencoder based on quantisation to generate such variables [Gendron et al., 2023c]. The next step will apply these concepts to architectures for reasoning tasks. As an application of our work, we have employed causal graph modelling to the auxiliary problem of animal behaviour recognition [Gendron et al., 2023b], described in Figure 2a.

3 Research Directions

We have investigated improvements to the LLM architecture to resemble causal mechanisms and increase their abstract reasoning abilities. Our future work in this direction will focus on regularising the training objective of LLMs to yield text generation that follows a high-level causal flow of thoughts. We will force the LLM generative process to write down causes before their expected effects (independently of their position in the output text) and discard spurious correlations to perform reasoning on unbiased causal variables. Finally, we will extend our variational graph architecture to solve abstract reasoning problems for vision. We will propose an architecture combining a variational autoencoder with a Bayesian neural network to answer unbiased causal intervention queries.