# Explainable Bayesian Networks via Natural Language Explanations and Interactive Visualisation

**Miruna Clinciu, Herbert Lau and Michael John Williams**
**Schlumberger Cambridge Research, Cambridge, UK**

## Overview

Bayesian Networks represent an important modeling technique that can deal with **uncertainty** in knowledge-based systems.

However, complex BNs draw diffidence and criticism due to insufficient transparency and information on how the results were reached.

**Key Takeaways:**

- Explainability
- Building **Trust** for expert and non-expert users
- Evaluation of **NL Explanations**
- Increasing Transparency in an **industrial setting**

## METHODS

### Interactive Visualisation

We added three "new ingredients" to a simple acyclic graph (DAG):

1. A specific colour code for inferred nodes
2. For each edge between nodes we added an infographic icon that will provide more information
3. Each node will receive a shape similar to those of a process flowchart diagram, therefore familiar to many engineering disciplines.

### NL Explanations

We provided three types of NL Explanations, produced by expert annotators, taking into consideration the technical aspects.

1. Causal NL Explanations (counterfactuals)
2. An explanation that provides an overall explanation of how a decision was achieved.
3. NL Explanations are attached to each edge, that will provide a technical reason.

Schlumberger

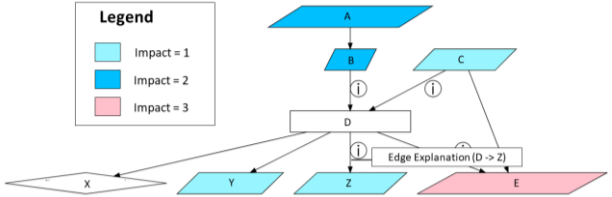


Figure 1. Interactive visualisation of a Bayesian Network, where colours indicate the relative importance of several key influences on the final decision; shapes are similar to those of a process flowchart (parallelogram: Input, rectangle: inferred and diamond :Output to decision)

## FUTURE WORK

In a study, these NL Explanations and different visualisation techniques will be rated in terms of informativeness, clarity, effectiveness and trust.