# Article information:

Methods for interpreting and understanding deep neural networks - ScienceDirect
<https://www.sciencedirect.com/science/article/pii/S1051200417302385>

# Article summary:

1. This paper provides an entry point to the problem of interpreting a deep neural network model and explaining its predictions.

2. It discusses techniques for interpreting complex machine learning models, with a focus on deep neural networks (DNN).

3. It distinguishes between post-hoc interpretability and incorporating interpretability directly into the structure of the model, and defines interpretation and explanation as two distinct concepts.

# Article rating:

Appears well balanced: The article presents the information in a reliable and balanced way, without biases and prejudices. The claims made in the article are well supported and, where applicable, all sides of the argument are given opportunity to present their point of view. The article appears trustworthy and reliable.

# Article analysis:

The article “Methods for Interpreting and Understanding Deep Neural Networks” is a comprehensive overview of techniques for interpreting complex machine learning models, with a focus on deep neural networks (DNN). The article is well-written and provides clear definitions of key terms such as “interpretation”, “explanation”, “post-hoc interpretability”, etc., which are essential for understanding the topic. The article also provides an overview of existing approaches to interpretation and explanation, as well as discussing recent developments in this field.

The article is written in an objective manner without any bias or promotional content. All claims made are supported by evidence from relevant sources such as research papers or studies. The article does not present any one side more than the other; instead it presents both sides equally and allows readers to draw their own conclusions based on the evidence presented. Furthermore, all potential risks associated with using these techniques are noted in the article.

In conclusion, this article is reliable and trustworthy due to its objective approach to presenting both sides of the argument equally without any bias or promotional content.

# Topics for further research:

* Deep Neural Network Interpretation
* Post-Hoc Interpretability Techniques
* Explainable Artificial Intelligence
* Interpretable Machine Learning
* Model-Agnostic Interpretation
* Interpretable Deep Learning Models

# Report location:

<https://www.fullpicture.app/item/22877eaa4ea51244ec2144729c4b9ed7>