LUNDS UNIVERSITET
Ekonomihögskolan
*Institutionen för informatik*

# Using Opaque AI for Smart Grids

## How does the difficulty of interpreting deep learning systems affect their suitability for smart grids?

Kandidatuppsats 15 hp, kurs SYSK16 i Informationssystem

Författare: Jesper Lundberg
Alexander Lundborg

Handledare: Blerim Emruli

Rättande lärare: Bo Andersson
Osama Mansour

# Using Opaque AI for Smart Grids: How does the difficulty of interpreting deep learning systems affect their suitability for smart grids?

ENGELSK TITEL: Using Opaque AI for Smart Grids: How does the difficulty of interpreting deep learning systems affect their suitability for smart grids?

FÖRFATTARE: Jesper Lundberg och Alexander Lundborg

UTGIVARE: Institutionen för informatik, Ekonomihögskolan, Lunds universitet

EXAMINATOR: Christina Keller, Professor

FRAMLAGD: maj, 2020

DOKUMENTTYP: Kandidatuppsats

ANTAL SIDOR: 74

NYCKELORD: Deep Learning, Smart Grids, Interpretability, Black-Box, Transparency, Post-hoc Explainability, AI, Machine Learning

SAMMANFATTNING (MAX. 200 ORD):
Deep learning is an emerging machine learning technique which can find complex patterns in large amounts of data. This makes it useful for several applications in smart grids, which often involve the processing of large amounts of data. However, there are reasons to be sceptical of its suitability as the black-box nature of deep learning could be a problem since power grids are important infrastructure and contain deadly currents. Professionals in smart grids were interviewed to provide an understanding of the importance of eight issues relating to the interpretability of machine learning. The findings show that for some uses related to controlling the grid, trust is of critical importance, and it is unlikely that a black-box algorithm will be used. For other uses such as giving recommendations and forecasts, it was found that either trust or informativeness is required for the results to be useful, although trust could potentially be achieved through a strong track-record, rather than through the ability to interpret the system. Other issues were of varying importance, but none of them critical.

Unless the area of interpretability sees considerable progress, it will be of concern when creating deep learning systems for smart grids.

# Contents

# Figures

# 1 Introduction

## 1.1 Background

According to a report by McKinsey and Co, Artificial Intelligence (AI) is poised to have a big impact on the economy in the coming years, with the potential to add trillions of dollars annually to the global economy (Chui, Manyika, Miremadi, Henke, Chung & Nel, 2018). AI is an umbrella term, covering several different techniques which are used for different purposes. The current wave of increased attention given to AI is fueled by advancements in machine learning, which are statistical algorithms that use an abundance of data to learn for themselves how to solve a problem. Although the definition of AI varies, machine learning is most often considered a part of it. The applications for machine learning include, but are not limited to image classification, voice recognition, reducing the energy use of data centres (Evans & Gao, 2016), and playing video games (LeCun, Bengio & Hinton, 2015). While machine learning has existed for a long time, recent advances in algorithms, data collection, and GPUs, have made machine learning stronger and more useful in practice (Zhang, Han & Deng, 2018). This paper concerns the application of these advancements in AI on real-world problems. The area of smart grids was selected, where the literature indicated that there seems to be a demand for recent machine learning techniques such as deep learning (Zhang, Han, & Deng, 2018; Shi, Xu and Li, 2018; Jindal, Aujla, Kumar, Prodan & Obaidat, 2018; He, Mendis and Wei, 2017). Smart grids can enable a larger proportion of electricity to come from renewable sources, adding to the importance of this area (Tuballa & Abundo, 2016). In this study, we have chosen to research the use of deep learning specifically.

### *1.1.1 Deep Learning*

Much of the improvement in performance for tasks such as image classification and game playing comes down to the use of *deep learning*, a machine learning technique that makes use of *neural networks* with a large number of so-called *neurons* (LeCun, Bengio & Hinton, 2015) which are arranged in layers. Note that it is the high number of layers made possible by modern computing that makes the network *deep*. While the concept of neural networks lends inspiration from biological brains, the neurons in neural networks are highly abstracted and simplified, but they both represent knowledge as connections in a network of neurons. Neural networks require large amounts of data but have the ability to find complex patterns and can often perform significantly better than other machine learning techniques at complicated tasks (LeCun, Bengio & Hinton, 2015; Shi, Xu & Li, 2017). One problem with deep learning, however, is that it is generally very difficult to interpret how the system arrives at its result, this is often referred to as the system being a *black-box*. Going back to the analogy of biological brains, interpreting deep learning systems by simply looking at the network would be similar to understanding human decision-making by opening up the decision-makers brain (Chakraborty, Tomsett, Raghavendra, Harborne, Alzantot, Cerutti, Srivastava, Preece, Julier, Rao, Kelley, Braines, Sensoy, Willis & Gurram, 2017). In contrast to humans, however, you cannot ask a deep learning system directly to explain the reasons behind their decision. While there is research on making the decisions more interpretable (often in a similar way to how humans explain their decisions) and some methods are already being used, these technologies

are still in their infancy and it would be imprudent to make assumptions about how successful they will be (these technologies are elaborated on in the literature review subsection on interpretability). The difficulty to interpret deep learning systems suggests that it could be challenging to use them in real-world settings, especially where there is an expectation of accountability.

### *1.1.2 Smart Grids*

A traditional power grid is a system that transmits electricity from generators to substations that transform the electricity to a higher voltage more suitable for transmission. These substations then transmit the electricity via transmission lines to other substations that transform it down to a lower voltage to finally distribute the electricity to consumers through distribution lines. This allows the electricity to be transported from the often remote areas where it is produced to the locations where it is needed.

There are many aspects that are considered part of a smart grid, but one of the main features is that power can be sent backwards, from customers to the grid. This has multiple benefits, one is that it makes home production of electricity from solar panels more economically viable since the electricity produced can be exported when it is not needed. Another benefit is that subgrids can function at a reduced capacity with locally produced power when disconnected from the main generators (Fang, Misra, Xue & Yang, 2012). Smart grids also include technologies that promote demand response when the supply is low relative to the demand, such as devices adjusting their energy use based on the status of smart meters (NIST, 2010). This means that a larger portion of the grid can be served by sources with uncontrollable or even unpredictable generation as the demand can be adjusted to fit the supply in real-time. Both solar and wind power, which are widely seen to be the main areas where renewable energy will grow, are examples of intermittent generation as the power generated is directly dependent on the weather. While hydroelectric is a renewable source of energy where the energy production can be controlled, the potential for further expansion is limited (IEA, 2010) and in 2019 the International Energy Association estimated both solar and wind power to expand considerably more than hydroelectric power between 2019 and 2024 (IEA, 2019).

As both smart and traditional power grids provide electricity to everyone, including important civic functions like hospitals, there is a necessity for reliability and accountability. Power grid operators in Sweden have legal responsibilities to ensure the stability and reliability of power transmission according to 3 kap. 9 § Ellag (SFS 1997:857). While electricity consumers are expected to be able to handle a 24-hour shortage, power grid companies are responsible for ensuring that no power outages are longer than that and cannot count on the electricity consumer having temporary solutions like back-up generators past that point (Energimyndigheten, 2009).

Furthermore, large parts of power grids contain extreme voltages that could easily kill a person. For this reason, it is of utmost importance that certain parts of the grid can reliably be turned off for maintenance of grid facilities or other work in proximity to high-voltage cables. The consequences if there would be a mistake could be fatal (Arbetsmiljöverket 2020).

## 1.2 Research Problem

Reliable electricity is a necessity for modern society, and the transition to carbon-free electricity is becoming increasingly important. In Sweden, a large coalition of political parties have joined an energy policy agreement with a target to phase out nuclear power by 2040 and have an energy production based completely on renewable sources (Regeringen, 2016). Since renewable energy sources like solar and wind power are intermittent, there is a need for a grid that can incorporate intermittent sources while maintaining reliability (Zhang, Han & Deng, 2018). Smart grids aim to solve this problem. However, smart grids require the analysis of large datasets to for instance forecast consumer behaviour for balancing supply and demand (i.e. load forecasting) (Zhang, Han, & Deng, 2018). According to Zhang, Han, & Deng (2018), deep learning, together with another modern AI technique called reinforcement learning, has the potential to meet these requirements. The data to feed into deep learning algorithms can come from for example smart meters. However, as mentioned before, the difficulty to interpret deep learning systems calls into question the suitability of deep learning for managing such an important piece of infrastructure as the electric grid. The decision made by the AI system cannot be easily understood, and this might be deemed necessary. The reason for this difficulty comes from the large number of neurons and self-generated connections in the deep neural network, the behaviour of which is too complex to interpret. This is why it is referred to as a black-box (Shwartz-Ziv & Tishby, 2017).

## 1.3 Aims

In this study, we want to explore how the difficulty of interpreting deep learning systems, because of their black-box nature, affects the possibility of using deep learning for smart grids.

## 1.4 Research Question

How does the difficulty of interpreting deep learning systems affect their suitability for smart grids in Sweden?

## 1.5 Delimitations

As mentioned, we limit ourselves to machine learning methods that use a deep neural network, i.e. deep learning. Also, while technology for aiding interpretability is briefly discussed, the study will not rely on potential future advancements, but rather what is currently possible.

# 2 Literature Review

## 2.1 Technical Overview of Deep Learning

As previously stated, a neural network consists of layers of neurons. Each neuron in one layer is connected to every neuron in the previous and following layer. The first layer of neurons is fed with the input data and this data then gets sent through each layer until an output is provided by the last layer (Nielsen, 2015). For instance, a picture of a handwritten digit can be given to a trained neural network. This picture will be processed by the network and finally a neuron in the last layer that corresponds to the correct digit will activate more strongly than the other neurons in that layer, indicating that the network "knows" what digit it has been shown. Knowledge is represented in the network by different strengths in the connection between neurons. This strength is called the weighting factor or simply *weight.* Additionally, a number called the *bias* is assigned to each neuron in order to adjust them correctly. The weights, biases and activations in the previous layer determine how strongly neurons activate, which is represented by a number from zero to one. Specifically, the activation of each previous neuron is multiplied with the weight of the connection to it and the bias is added, this is then put through a function that compresses the number to somewhere between zero and one (Nielsen, 2015). To get their knowledge, the neural network has to learn from data. This is done by calibrating the initially random weights and biases through a method called *backpropagation,* which is done automatically. This method is complicated so here we provide a simplified and incomplete example where we follow the steps manually. The example focuses on calibrating the weights and we will only be looking at how the weights should be increased.

The graph below (figure 1.1) illustrates an example of backpropagation in a shallow neural network, which is untrained and therefore currently gives the wrong answer. In the example, we want the neuron corresponding to the correct answer, in this case "2", to get a higher activation (a higher number in the neuron marked as $C_2$ in the graph). We achieve this by slightly increasing the weights of the connections (shown with bold lines) from the highly activated neurons in the previous layer ($B_1$ and $B_2$). This is the most effective way of increasing the output neuron's activation since the activation of the previous neuron and the associated weight are multiplied (Sanderson, 2017a). In the next step we move on to the previous layer, which is the second to last layer (in this case the middle layer) and we want to find the most valuable neurons in this layer and increase the weights towards them. The most valuable neuron in the second to last layer is the neuron with the highest weight towards the correct answer (signified by the numbers from zero to one on the connections), in this case it is the neuron noted with $B_3$ that is the most valuable. We therefore increase the weights of the connection from the neurons with the highest activation in the leftmost layer ($A_1$ and $A_2$) to the most valuable neuron in the second to last layer ($B_3$). In deep learning networks, there are many more layers and a high number of neurons in each layer. Weights would also be lowered in a similar way to decrease the activation of the wrong answers and changes are not binary, you do not determine if a weight should be increased, but rather, how much and in which direction it should change. This process can be repeated on any number of layers by seeing how the neurons affect valuable neurons in the layer after it.

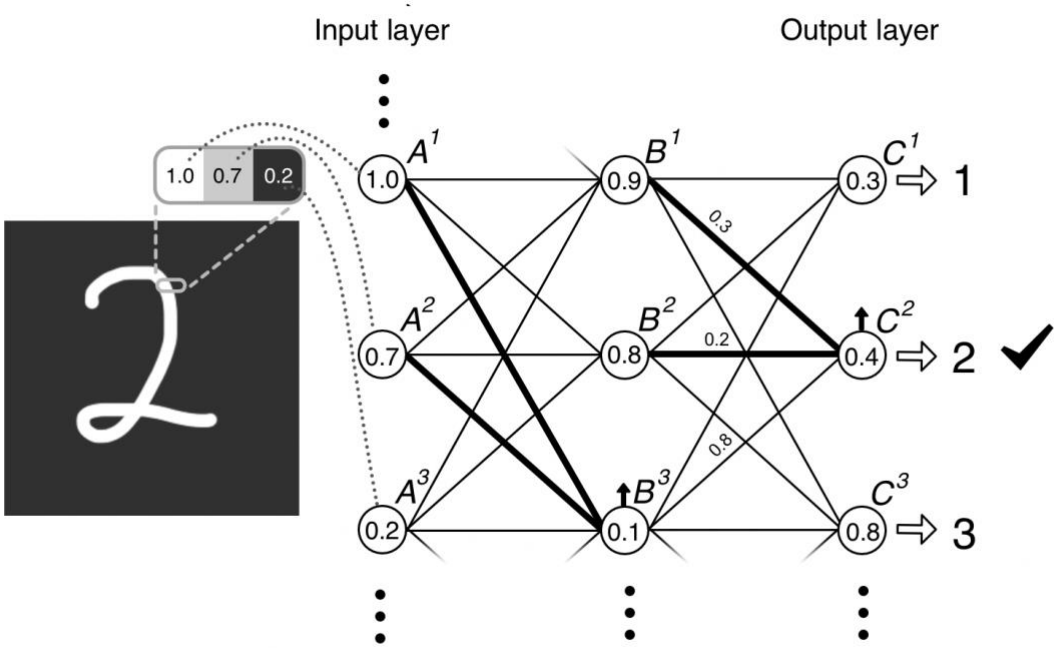


**Figure 1.1:** An illustration of a shallow neural network that is learning from a picture of a handwritten two. The bold lines mean that the weight of those connections should increase, and the up arrows indicate that those neurons will get a higher activation once the weights have been adjusted.

After doing this with every picture in the training data on every layer, the neural network will learn to more accurately select the correct output (Sanderson, 2017a). This entire process can then be repeated to improve the performance even further until it is almost as good as it can get with this data set and further improvements are negligible. This will make the neural network able to recognize very complex patterns by having the neurons represent certain elements. In the example with number recognition, it can be likened to the neurons representing different building blocks of numbers, such as lines and loops, but in reality, the patterns that are found are beyond human understanding and look like nonsense (Sanderson, 2017b). It bears reminding that this is a simplified version and real systems are usually more complicated. It should also be noted that unless the dataset is sufficiently large relative to the size of the neural network, the results will be too specific to be generalisable to new data, this problem is called “overfitting”.

## 2.2 Application of Deep Learning on Smart Grids

We have identified three promising applications for deep learning that applies to smart grids: load forecasting, demand response and false data injection detection.

### *2.2.1 Load Forecasting*

Load forecasting refers to the prediction of future power consumption, which is important for the planning of power generation. It could be a matter of determining if production needs to be increased in a few hours or whether an investment will pay off in a few years (Anwar, Sharma, Chakraborty & Sirohia, 2018). According to Shi, Xu and Li (2018), individual

household electricity use displays a high amount of uncertainty and most systems for predicting short term electricity use are ill-equipped to deal with this level of detail. However, according to Shi, Xu and Li (2018), this can be overcome by deep learning as it can account for many of the external factors that cause the uncertainty in a way that traditional systems would be unable to. Furthermore, deep learning could find out how these factors affect each individual household. An example Shi, Xu and Li (2018) bring up is that sunlight is likely to cause a larger reduction in power demand for a household with solar panels installed. An implementation of deep learning is proposed, and while Shi, Xu and Li (2018) advise caution in reading too much into the result, the deep learning implementation significantly outperforms traditional methods in their simulation, which is a promising sign for the use of deep learning for load forecasting. The increase in solar and wind power generation in electric grids also brings uncertainty due to their intermittency which increases the challenge of balancing supply and demand. Zhang, Han and Deng (2018) explain that traditional methods of load forecasting are not able to deal with the large quantity of data that is received from smart meters, devices that measure the consumption of electricity in a household. Similarly to Shi, Xu and Li (2018), they find that deep learning systems can deal with the large quantity of data and are useful for predicting power load on both a household and aggregate level.

### *2.2.2 Demand Response*

Demand response is a way to balance demand and supply of electricity by incentivising consumers to adjust their electricity usage so that less is used at the peak of demand (Albadi & El-Saadany, 2008). Lu and Hong (2019) discusses two types of demand response: price-based and incentive-based. An example of price-based demand response is for a customer to set the dishwasher to run its program during a time when electricity demand is low in order to get a reduced electricity bill. Deep learning systems can be used to calculate when consumers should use electricity to get the lowest price and reduce the demand-peaks (Zhang, Han & Deng, 2018). The other type, incentive-based, rewards customers for using less electricity and is the most used demand response method in the US according to Lu and Hong (2019). Deep learning has been shown to find hidden patterns in electricity consumption which is useful for demand response programs (Jindal, Aujla, Kumar, Prodan & Obaidat, 2018). Lu and Hong (2019) proposed a system based on deep learning to enable incentive-based demand response which they claim can improve the reliability of the grid and improve provider and consumer profitability, based on results from a simulation. Ruelens, Claessens, Vrancx, Spiessens and Deconinck (2019) studied a demand response situation where a residential heating system and water heater are controlled to optimize its electricity use. They show that a long short-term memory network, a deep learning architecture with internal memory, achieves better results than other deep learning techniques.

### *2.2.3 False Data Injection Detection*

*False data injection* (FDI) attacks are a type of attack based on giving the power grid false data first theorized by Yao, Peng and Reiter (2011). Because of the reliance on smart meters, which are often lacking in security, smart grids are more susceptible to FDI attacks than traditional power grids. (Lin, Yu, Yang, Xu & Zhao, 2012)

He, Mendis and Wei (2017) present a method of using deep learning to detect FDI attacks, in this case for the purposes of stealing electricity. Through simulation, they show that deep learning can be used to great effect for identifying which data is forged. It is, however,

important to note that He, Mendis and Wei (2017) had to create more false data based on some existing false data. This is because there is a lot more data for the system working normally than there is for false data injections, this is an issue when deep learning solves binary classification problems (problems where you decide which of two categories a set of data belongs to) called unbalanced data and it will make the neural network overly biased towards picking the category for which there is more data.

Niu, Li, Sun and Tomsovic (2019) claim that this type of simulation gives a misleading sense of the detection giving less false negatives than the system actually would have in reality, because real-world data is unbalanced. However, they never explain why a real-world system could not also create simulated false data. Instead, they present a system that learns how the grid normally behaves and identifies anomalies without relying on false data. This also makes it able to detect attacks which do not follow an earlier pattern. Regardless of which is the best approach, it seems clear that deep learning is useful for identifying false data injection attacks.

## 2.3 Black-box Challenges

Due to the large number of neurons and connections in real-world neural networks, and since the knowledge stored in them is self-generated, it is very difficult to understand how they derive their results. As mentioned, this makes them black-boxes (Shwartz-Ziv & Tishby, 2017), which are systems where an input leads to an output but the process by which this is done is hidden (Bunge, 1963). This can lead to challenges when applying deep learning to smart grids.

### *2.3.1 Interpretability*

In recent years there has been a lot of research into making machine learning easier to comprehend, the property that expresses the possibility of comprehending a system is usually referred to as *interpretability*. The word explainability is sometimes used instead, however, the definition of explainability varies and sometimes explainability does not cover all the relevant concepts (namely transparency which will be introduced below) and therefore interpretability will be used throughout this study. The research on this subject generally deals with broader concepts like machine learning or AI, which nonetheless encompass deep learning.

#### *2.3.1.1 Goals of Interpretability*

Through analysis of earlier sources' use of the term, Lipton (2016) finds that interpretability tends to be desired for five purposes. One is *trust*, meaning that the system can be trusted to achieve its objective better than current methods without introducing new problems. The second is *causality*, meaning that you can see which correlations found by the system are causal. That is, those correlations that consist of a cause and an effect. The third is *transferability*, meaning that what has been learned by the system can be transferred to a similar problem. The fourth is *informativeness*, if the purpose of the system is to provide information to help with human decision making, it is useful to know how the conclusions have been reached, just as this article should include more than the conclusion. Finally, Lipton (2016) identifies *fair and ethical decision-making*. Machine learning can often reproduce human biases or find unethical ways to exploit people (for example marketing alcohol to

alcoholists), interpretability has been discussed as a way to identify and intervene in these patterns.

Arrieta, Díaz-Rodríguez, Del Ser, Bennetot, Tabik, Barbado, Garcia, Gil-Lopez, Molina, Benjamins, Chatila, and Herrera (2020) reaffirms these purposes for interpretability (although fair and ethical decision making is replaced with fairness and trust is called trustworthiness) and adds *confidence*, *accessibility*, *interactivity,* and *privacy awareness*. Although the phrase "explainability" is used, it is quite clear that this is meant to build upon the findings of Lipton (2016), and it is therefore reasonable to assume that it is used as a synonym to interpretability. Additionally, Arrieta et al. (2020) does not present any significant distinction between trust and confidence, and as such, they will be treated as the same thing in this article. Accessibility means that people who are not experts, like end-users, can understand the system and its decisions. Interactivity means that end-users are able to interact with the system. It is not quite clear what type of interaction this would be, but presumably the idea is that all kinds of interactions are made easier if you know how the system works. The examples from the references in Arrieta et al. (2020) is military robots reporting to a commander who need to explain their decisions (Langley, Meadows, Sridharan & Choi, 2017) and training systems where learning can be enhanced by knowing how virtual actors made their decisions (Harbers, van den Bosch & Meyer, 2010). Privacy awareness is important as the deep learning system may find patterns that constitute a breach of privacy. Even if the input data is not a violation of privacy, it is possible that a conclusion drawn by the system is. An example of this is that although names are not in the special categories of data that GDPR provides extra protection for, it is easy to imagine a machine learning system using names and other normal forms of data to draw conclusions about people's ethnicity, which is one of these special categories of data. Similar things could happen with using salary, vacation time, etc. to figure out someone's union membership or an advertising bot using political views as a step in finding out the best way to advertise based on social media activity. With a lack of interpretability, this could happen in a deep learning system without the developers' knowledge or intent.

#### *2.3.1.2 Categories of Interpretability*

Based on these goals of interpretability Lipton (2016) presents a number of categories of interpretability, grouped into *transparency* and *post-hoc explanations*. Transparency means that a person can understand how the system operates and the concept is composed of *simulatability*, which means that the system is simple enough for a person to understand as a whole, *decomposability*, meaning that a person can understand the individual components of the system, and *algorithmic transparency*, meaning that people can understand the underlying algorithms that create the system. Post-hoc explanations are explanations that, rather than rely on making the inner workings of the system understandable, provide explanations for the reached conclusions in a way that is suitable for human understanding, Lipton (2016) points out, however, that this often overlaps with transparency as the explanations tend to be based on the state of the system. Lipton (2016) presents four different ways of doing this. The first is *text explanations*, which could, for example, be achieved by creating a separate machine learning system for the purpose of explaining the original. The second is *visualization*, which means that you try to make the system more understandable by visualizing it. The third, *local explanation*, means that you explain part of the system's behaviour, for example, you might highlight parts of a processed image or visual representation of the problem that is most important to determine the output of the system. The fourth is *explanation by example*, which means that the system provides examples of similar cases in the training data. This can be compared to how people use analogies to explain their behaviour.

Arrieta et al. (2020) acknowledge the aforementioned division of transparency and post-hoc explanations and like Lipton (2016) divides transparency into simulatability, decomposability, and algorithmic transparency. For post-hoc explanations, Arriata et al. (2020) confirms the four types of explanations from Lipton (2016), adds *explanations by simplification*, and separates local explanations into local explanations and *feature relevance explanations*. Explanations by simplification means that you generate a simpler algorithm that produces a similar result to the original algorithm and feature relevance explanations means that you point out the relevance of different variables for affecting the output while local explanations means that you explain part of the system. As the degree of transparency is inherent to an algorithm and deep learning algorithms are opaque (interpretability applies to all machine learning, not just deep learning), they need to rely on post-hoc explanations. Fan, Xioung, and Wang (2020) does not explicitly adhere to the separation of transparency and post-hoc explanations but rather describes interpretability to consist of the three parts that the other articles describe as transparency, while what others describe as post-hoc explanations are instead described as "interpretation methods". This is not as incompatible as it might originally seem, since, as previously mentioned, Lipton (2016) describes that post-hoc explanations often try to aid the type of interpretation that transparency entails. Many of the methods that Fan, Xioung, and Wang (2020) present directly match specific types of post-hoc explanation. Proxy is essentially the same as explanation by simplification, saliency the same as feature relevance explanation, feature analysis the same as local explanations, explaining-by-text the same as text explanations and explaining-by-case the same as explanation by example. However, they also present two previously unmentioned methods, Advanced mathematics/physics and model inspection. Advanced mathematics/physics refers to the use of advanced maths or physics to describe systems or their outputs and model inspection seems to essentially be a "none of the above" category, with a focus on interpreting the system as a whole as opposed to certain aspects of it.

### *2.3.2 Debugging, Testing & Validation*

The black-box nature of deep learning systems, including the fact that they are largely self-generated makes them difficult to debug. According to Odena and Goodfellow (2018), it takes a lot of computation to obtain basic information about a neural network. Another reason why they are difficult to debug is that they often differ in practice from theoretical models of neural networks in which it is possible to verify their properties. Odena and Goodfellow (2018) introduced testing techniques for deep learning that can be used to find errors that only show up for rare inputs. Sun, Huang, Kroening, Sharp, Hill and Ashmore (2019) proposes another testing and debugging tool, which they claim is rigorous enough to be used in safety-related applications and which can be used to evaluate the internal workings of deep neural networks.

Verifying the behaviour of deep learning systems in safety-critical applications is both important and difficult. Zhao, Banks, Sharp, Robu, Flynn, Fisher and Huang (2020) note that "the performance and explainability of [machine learning] models within practical critical systems require a rigorous and continuous verification of their safe utilisation". A failure in a deep learning system used in smart grid applications can have serious consequences like large power outages and it is, therefore, important to rigorously test and validate them. This is related to the previously described concept of trust in the context of interpretability, verifying a system's behaviour through testing and validation is a way of attaining trust. A well-known safety issue for deep learning systems are *adversarial examples*, slightly and possibly

undetectably altered inputs that cause the system to output a completely different output (Sun, Huang, Kroening, Sharp, Hill & Ashmore, 2018). This has been discussed with regards to self-driving cars where the system controlling the car is "fooled" by signs which have been slightly altered (Sitawarin, Bhagoji, Mosenia, Chiang & Mittal, 2018). Whether this could pose a problem in the context of smart grids is something we have not been able to discern.

## 2.4 Summary of Literature Review

In this chapter, we presented an in-depth explanation of deep learning, we covered its structure and the backpropagation method used for making the system learn from data. We then mentioned three applications of deep learning on smart grids, namely load forecasting, demand response and false data injection detection. The first one, load forecasting, comprises the use of deep learning to predict the future demand for electricity so that the grid can balance demand with supply. The second one, demand response, comprises different ways to use deep learning to adjust demand and avoid high demand peaks, for instance by motivating customers to postpone electricity use to a period which is found to be optimal for price reduction and smoothing out the demand curve, or controlling building heaters and water heaters directly. Thirdly, false data injection detection is the process of finding and removing false data from e.g. smart meters.

We also discussed several challenges with implementing and using deep learning. We found that challenges arise from the nature of deep neural networks. In particular, their inner workings are hard to interpret, they are said to be so-called black-boxes. This gives rise to problems when there is a need to know why a certain result or decision is made. It also gives rise to challenges during testing, debugging, and validation since it is difficult to understand why an error has occurred and ensuring that the system works as intended for all possible data inputs. We brought up the concept of interpretability, which is the ability to understand how an AI system works. Interpretability can be broken up into transparency and post-hoc explanations. Transparency refers to the ability to understand how a system works by studying it. Transparency consists of three aspects, simulatability, decomposability, and algorithmic transparency. Simulatability means that the system is simple enough to be understood as a whole, decomposability means that the individual components of the system can be understood, and algorithmic transparency refers to the knowledge to understand the underlying algorithm that created the system. Post-hoc explanations are methods for understanding the reasons for the system's result by analysing the system indirectly, e.g. by using a different system to provide explanations or highlighting the part of the input that was most important for the system to arrive at its conclusions. We have also discussed eight advantages of interpretable AI systems: trust, causality, transferability, informativeness, fair and ethical decision-making, accessibility, interactivity, and privacy awareness. The black-box nature of deep learning makes transparency unfeasible, so to understand a deep learning system one must rely on post-hoc methods. To conclude, we have found several applications of deep learning for smart grids but have also found challenges with implementing and using it due to the difficulty of interpreting the inner workings of a deep learning system.

# 3 Methodology

## 3.1 Choice of Methodology

Information was collected through the use of Google Scholar, using search queries such as "allintitle: 'deep learning' 'smart grid'", "'load forecasting' 'deep learning'", "'deep learning' debugging" et cetera. We prioritised articles with many references and which had been peer-reviewed since it is a sign that the article is more reliable and of higher quality, but when necessary we also read working papers and articles submitted to conferences. We also read articles that referenced the articles we found when searching in order to read other authors' responses to the original articles as well as articles that the original article referenced. For the collection of empirical material, we chose the qualitative method of doing semi-structured interviews with individuals with relevant roles that make them informed about smart grids and possibly experienced with applying deep learning on smart grids. Based on the research question, which requires detailed knowledge about the domain of smart grids in order to answer, it is fitting to interview professionals at organisations with a connection to smart grids. Interviews were preferred over surveys and literature reviews since there is a need for rich knowledge about not only the business of smart grids, but also the culture surrounding it.

## 3.2 Choice of Interviewees

In the choice of interviewees, we decided to cast a wide net since we expect that the object of study, the implementation of deep learning on smart grids, is a novel and niche subject with a limited number of relevant potential interviewees. We sent requests to several companies that we found through internet searches and had reason to think could have the relevant experience and expertise, informed them of our purpose and asked to interview a person at the company with experience of smart grids and specified that we could be interested in interviewing someone even if they do not have any particular experience of deep learning. We specified this so that we would not exclude interesting potential interviewees by being too selective.

## 3.3 Research Quality

For qualitative studies such as this one, it is important to conduct the research with validity and reliability in mind (Jacobsen, 2002). The concept of validity is composed of internal and external validity. Internal validity refers to the validity of the results. The validity of the results is strengthened if multiple people agree with them or if other studies reach similar results using different research approaches. External validity refers to the degree to which the findings generalize. For qualitative research, this usually means to generalise from the collected data to theory in contrast to quantitative research where the aim is often to generalize to a larger population of data samples.

Reliability, on the other hand, is about the research process. It is in practice virtually impossible to conduct research without affecting the subjects to be researched in some way. According to Jacobsen (2002), the research subjects can be affected by the environment that the subjects are in. They are divided into artificial and natural, where artificial environments, or contexts, are environments that are unusual for the subjects, for instance the researcher's office or in a laboratory. Natural contexts, on the other hand, are environments that the subjects are used to, for instance at home or in their workplace. Research shows that the environment that people are situated in affects their behaviour, and for this reason, many researchers prefer to conduct their research in natural contexts (Jacobsen, 2002). In our case, our interviewees are interviewed in their workplaces, at a date and time chosen by them and where they expressed a preference for using a certain teleconferencing platform, it was used if possible. This means that we have taken steps to ensure that our research is conducted in a natural context. To avoid inaccurate representation of the information from the interviews, all interviews were recorded with the interviewees' permission.

## 3.4 Research Ethics

Oates (2006) brings up five rights of participants in research, namely: right not to participate, right to withdraw, right to give informed consent, right to anonymity and right to confidentiality. The first right says that a person who does not want to participate in research does not have to and should not be forced to participate (Oates, 2006). Secondly, a participant has the right to withdraw from the research at any time. This includes the right to opt-out of certain parts of the research, by for example not answering a particular question (Oates, 2006). Thirdly, a participant's consent is only given once they have been provided with the complete nature of the research and their role in it. The researchers should also make the participants aware that they have a right not to participate and that they can withdraw at any time (Oates, 2006). Fourth, participants have the right to have their identity and location protected and disguised if there is a need for it, for instance with pseudonyms. Organisations' identities should also not be revealed except if they explicitly allow it (Oates, 2006). The last one is the right to confidentiality. Participants have the right to have the data that has been collected from them be confidential if they want to, this means for example that a researcher should not leave the data in public where anyone can look at it. Additionally, if a participant tells the researcher something in confidence it should of course not be included in the research report (Oates, 2006). We have made sure to respect these rights when conducting the interviews.

## 3.5 Implementation

Before each interview, we explained the purpose of the study. We also informed the interviewees that the result will only be used for the purposes of this study, that the interview is voluntary and that the interviewee can terminate the interview at any moment. We also informed them that the interview will be recorded and asked them if they wanted to be anonymized. Since we conducted the interviews with Swedish companies, we chose to use Swedish as the language for the interviews and we have thus written the interview questions in Swedish, but we have included an English translation in the appendix, along with the original Swedish questions.

## 3.6 Interview Guide

The guide for our interviews is informed by our literature review and are intended to shed light on our research question. To avoid making invalid assumptions, we decided to not select questions about issues with black-box systems based on relevance to smart grids. In some cases, the terminology was simplified to make the questions easier for the interviewee to answer. All questions were sent at least 24 hours in advance to each interviewee to give them time to prepare for the interview. However, we did not send all the prepared follow-up questions to prevent the interviewee from being influenced by them when answering the other questions. We started the interview with some background questions to get information about the interviewee's role and experience. We subsequently asked them whether their company is using deep learning for smart grids, if they have experience implementing deep learning and if they see any appropriate application of deep learning for smart grids. We went on to ask them what challenges they have experienced or expect when applying deep learning for smart grids. We note that a major problem with deep learning is the difficulty to understand how it arrives at its conclusions, we therefore asked how this affects the possibility of using deep learning for smart grids. We found eight advantages in the literature of interpretable deep learning systems, so we also asked them about the relevance of these advantages to their smart grid business, in order to assess how problematic their absence would be in a black-box system. Our final question was whether the interviewee could think of anything else that is relevant to our research which hasn't been brought up yet. We then thanked the interviewee for participating and finished the conversation.

# 4 Results

Here we present the results from our interviews. We interviewed professionals from four organisations: Vattenfall Eldistribution, Ellevio, Mälarenergi and Energimyndigeten. The first three of which are electricity network operators, which means that they distribute the electricity that customers buy from electricity suppliers. Vattenfall Eldistribution and Ellevio are two of the three main companies operating regional grids (Miljödepartementet, 2007). Vattenfall Eldistribution is a subsidiary of the government owned power company Vattenfall (Vattenfall Eldistribution, n.d.). They have about 900 000 customers connected to their grid. Ellevio has 960 000 customers, most of whom live in Stockholm (Ellevio, n.d.). Mälarenergi Elnät AB has about 150 000 customers and is based in Västerås (Mälarenergi, n.d.). Energimyndigheten is the Swedish Energy Agency, which is sorted under the Infrastructure Department. They support research on smart grids, among other things (Energimyndigheten, 2020).

## 4.1 Areas of Use for Deep Learning in Smart Grids

Vattenfall Eldistribution currently uses machine learning for data analysis and pattern recognition of e.g. data from smart meters to make forecasts. However, it was not clear from the interview if they use deep learning specifically. According to the interviewee, machine learning is most useful for short forecasts and becomes less accurate for longer time spans. Ellevio currently uses machine learning to make forecasts of expected demand, although this is not a smart grid application specifically. The interviewee expected deep learning to be useful when automating monitoring and remote operation of the grid. The interviewees from Mälarenergi brought up that the main area of use was to create new load profiles which are more up-to-date as the current ones were made in the ’70s. The interviewees from Energimyndigheten saw many applications of deep learning for smart grids, one of which would be long-term forecasts to build infrastructure in order to prevent capacity shortages. Power grid companies are also looking into using large amounts of data for predictive maintenance, but there does not seem to be any plans for using deep learning. There are also projects for using deep learning to find the reasons for disturbances near solar and wind power plants. In general, the growth of intermittent and uncontrollable power sources and microproducers has increased the need for new solutions that often need to handle large amounts of data. Some type of advanced algorithms will probably be required for future energy markets, which include aggregators and demand response, but it is still unclear how that will happen.

## 4.2 General Challenges

According to the interviewee from Vattenfall Eldistribution, the business of power grids is conservative and new methods, in general, need to gain trust. They added that this trust is gained by the algorithm giving consistently good results over a long period of time. The

interviewee suspects that gaining this trust might take longer than developing the technology. The interviewee from Ellevio stated that a lack of measurement data can pose a challenge. The interviewees from Mälarenergi said that for the purpose of creating demand profiles it seems more economical to simply prepare for higher loads than to create advanced systems for precisely predicting power use. Especially since their infrastructure is meant to last for decades which means power use will be unreliable either way. For demand response, they thought that there are too many actors involved and as for grid control the risks are too high to trust a machine, as there are repairmen on location when something goes wrong who could be electrocuted if the grid is not properly managed. Most of the grid is not smart which means there is not a lot of data and control points for a deep learning system to train on or control. Furthermore, the grid needs to be reliable, so if the deep learning system shuts down, you need redundancy. Energimyndigheten said that one challenge is to aggregate the data in order to anonymise it, and even then there are difficulties in keeping people anonymous even when data is anonymised, which is elaborated on under privacy awareness. Another challenge is to make different systems interoperate with each other. Further, they brought up that it is not clear who should do everything, for example, it is not clear who should pay for smart meters as multiple stakeholders have an interest in seeing them installed. But the biggest obstacle according to the interviewees from Energimyndigheten is that the system has to be trustworthy and reliable.

## 4.3 Black-box Challenges Brought Up by Interviewees

The interviewee from Vattenfall Eldistribution said that while trust, as they mentioned previously, has to primarily be gained through a strong track record, interpretability can expedite the process. The interviewee also said that there are issues with understanding current models. Currently, when managing Ellevio's systems it is important to know exactly why something is done. The interviewee can imagine that operations could be made automatically in the future without anyone understanding why they were made in a particular way and in that case, there is a need for trustworthiness. The interviewees from Mälarenergi brought up both that you need to have rules for what the system can do because some actions could cause huge costs in being unable to distribute power properly. One of the interviewees from Energimyndigheten mentioned both issues with trust and informativeness (or in her words, traceability) as black-box challenges. They added that these issues were a larger concern in operating and maintaining the grid than in long term forecasting where it is not as important to understand how the deep learning system arrives at its conclusions if they seem reasonable and in that case, they can inform human decisions. There could be applications, according to the interviewees, in operating the grid where black-box systems could be accepted, for instance, if there is a risk of a big power outage and there is a need for quick decisions to be made where there is no time to check the reasons behind the decision. But in general, they thought that one of the big obstacles for using deep learning in power grids (probably referring to operating the grid) is that you have to trust that it would not cause blackouts, especially to certain vital civic services, like hospitals.

## 4.4 Relevance of Trust

Trust was deemed to be important by our interviewee from Vattenfall Eldistribution, and as mentioned, it can be gained through experience. Another important requirement for gaining

trust was the system's sensitivity to flawed input. They stated that a few incorrect data points should not result in a useless forecast. Trust, or trustworthiness, was stated to be very important by the interviewee from Ellevio since personal safety is their highest priority. Personnel can be working in the grid when remote operations are made and it therefore matters that operations are carried out safely, they said. The interviewees from Mälarenergi were, as previously mentioned, very sceptical about letting any AI control the grid, as mistakes could potentially cause workers to be electrocuted and die, meaning you would need a level of trust that the interviewees seemed to think is impossible to achieve from any automated system. However, for pure forecasting, the interviewees did not have any qualms over trust. It also came up that for the equipment used in these instances, very thorough tests have to be done. Trust was said by our interviewees from Energimyndigheten to be very important. It could, however, be earned without the ability to understand the system in some areas of use. It would, in that case, need to be tested thoroughly, they said. The greater the risk involved in the area of use, the greater is the need to understand the system. Risk is higher in operations and maintenance and low in analysis and forecasts and having a conceptual model of the system was also said to aid in developing trust.

## 4.5 Relevance of Causality

Due to miscommunication, we were not able to get an answer to the question as we intended it from our interviewee from Vattenfall Energidistribution. We unfortunately did not realize this until after the interview. Understanding the cause and effect relationship was said to be important by the interviewee from Ellevio, for example when dealing with electricity-quality problems. There is also a large number of factors involved that can affect the end result. The interviewees from Mälarenergi said that causality was potentially interesting but one of the interviewees was not sure if the data required to find cause and effect relationships would be available. One of the interviewees from Energimyndigheten thought that the question was complex and implied that it was difficult to answer but noted that information about weather and other information about world events can be useful.

## 4.6 Relevance of Transferability

The interviewee from Vattenfall Eldistribution thinks transferability will be useful in the future when deep learning is more widely used, but cannot think of any specific instances right now where transferability would be useful. The interviewee from Ellevio stated that there will be many situations where the solution used for one network operator's problem can be transferred to another one by a supplier of deep learning systems since the network operators use unique, but very similar methods. One of the interviewees from Mälarenergi could imagine that transferability could have advantages but admitted that the application they described was purely hypothetical. The other interviewee could also see uses for it. On a different question, one of the interviewees pointed out that when something goes wrong, power has to be rerouted, which changes the problem environment. One of the interviewees from Energimyndigheten said that it would be useful if the system can take on a new problem that is similar to previous problems it has solved but it was said to not be as important as trust.

## 4.7 Relevance of Informativeness

According to the interviewee from Vattenfall Eldistribution, the importance of reviewing decisions depends on how critical the result is. Furthermore, it is not an absolute requirement, but as previously mentioned it could rather speed up the process of gaining trust. If the system cannot be understood, more time is needed for the system to be trusted. The interviewee from Ellevio said that it is very important for them to have documentation of why something went wrong, for example, if a power outage happened, including why certain human and computer decisions were made. Even if the decision cannot be reviewed in real-time, this information is going to be very important for investigations after the fact. The interviewees from Mälarenergi deemed informativeness to be essential for safety functions and areas where stability is important but merely useful for other less critical applications where it could automate the work of figuring out why something occurred. This issue was very important according to the interviewees from Energimyndigheten, but like the interviewee from Ellevio, information about the decision does not always matter at the time of the decision but instead when following up the decision afterwards. In some situations, where a system recommends a decision but does not provide a reason, the recommendation can still be accepted if additional data reaffirms the decision. Also, if the system is highly trustworthy, this diminishes the need for understanding the reasons behind its decisions.

## 4.8 Relevance of Fair and Ethical Decision-Making

The interviewee from Vattenfall Eldistribution pointed out that while they have access to measurement and use data from customers, there are regulations in place, like unbundling to prevent abuse. The interviewee from Ellevio says that since most decisions are made between machines, there are not a lot of parallels that could be drawn to the examples of making decisions that directly affect individuals. However, the interviewee speculated that there could be issues where companies are treated unfairly. The interviewees from Mälarenergi were in disagreement about whether there was any potential for unethical results of using deep learning. One of them entertained a potential scenario where the AI system would favour customers who consume a lot of electricity over customers who consume less, but it was made clear that this was hypothetical. While the interviewees from Energimyndigheten said that while fair and ethical decision making was important in general, they did not think there was any potential for jeopardizing it with smart grids.

## 4.9 Relevance of Accessibility

The interviewee from Vattenfall Eldistribution pointed out that accessibility would help the system gain trust quicker among the parts of the organisation that are not experts. The interviewee further explains that this does not have to happen through explaining the algorithm itself, but rather through explanations of the more practical issues like what input gives what output. Accessibility is of great importance to Ellevio according to the interviewee as they rely on entrepreneurs who are out in the field who build, maintain, and operate the grid and cannot be expected to have expert knowledge on the system or deep learning in general. The interviewees from Mälarenergi said that accessibility was necessary for operation-critical applications because they wanted to avoid a dependency on experts in case something would go wrong. However, for non-critical analysis this was not necessary. This

issue was said to be potentially relevant by one of the interviewees from Energimyndigeten if employees in companies related to the grid would need to use artificial intelligence to aid them in their tasks.

## 4.10 Relevance of Interactivity

The interviewee from Vattenfall Eldistribution said that they want corporate customers to voluntarily give them their forecasts for energy use. For this to happen it has to be easy to interact with the system. The interviewee from Ellevio claims that this is a very important factor as they have to be able to cooperate with suppliers who need a simple interface and customers who need a very simple one. An example could be that there have been requests for a system that shows where the grid has capacity for more power production. Most of this was brought up on the question of accessibility since they are closely related concepts. The ability for end customers to interact with the deep learning system was said to not be necessary for Mälarenergi according to the interviewees. However, other companies such as electricity suppliers may want to give the customer advice from the system about when to use electricity to get a low price. Interactivity could be relevant when different network operators would have to interact with each other's system according to the interviewees from Energimyndigheten, but this issue was said to be less important than some of the others.

## 4.11 Relevance of Privacy Awareness

Vattenfall Eldistribution only uses personal information for the purposes of billing, which does not give rise to privacy violations. The data is also aggregated, which improves privacy. The interviewee from Ellevio explains that power usage can be used to deduce a lot of personal information, when someone goes to work, if they are at home sick, and in cases in the USA, it has even been used to detect plant lights used for secret marijuana plants. The interviewees from Mälarenergi said that data about customer electricity use had to be protected and kept private but they could not see any way that data about electricity use could pose threats to privacy. The interviewees from Energimyndigheten brought up privacy when asked about general challenges, one of the interviewees brought up that data has to be aggregated to protect privacy and the other interviewee responded that it might be possible to match aggregated data to individuals, bringing up as an analogy that keyboard inputs can be used to find patterns that can identify people. However, this is purely hypothetical and the interviewee did not know whether this was actually possible.

# 5 Discussion

One thing that we have found out during the interviews is that deep learning is not a well-known term among employees at companies that work with smart grids. They mostly have some idea about what machine learning is but we found ourselves having to explain deep learning to the interviewees. This state of affairs makes it harder to ascertain a conclusive answer to the larger question about the usefulness of deep learning for smart grids, but by asking about the eight advantages with interpretable systems, we can probe into whether deep learning without the augmentation of interpretability, can be utilized for smart grids. We will now go through each advantage of interpretability from Lipton (2016) and Arrieta et al. (2020) and discuss their perceived importance based on the result from our interviews.

## 5.1 Interpretability Issues

### *5.1.1 Trust*

It was clear from all interviews that trust is very important. However, it was not fully agreed whether trust requires interpretability or can be achieved solely through a strong track record. The interviewee from Vattenfall Eldistribution, who was the main proponent of the track record idea, claimed that you need to trust the system to be able to handle any type of invalid data without serious consequences. It is not immediately obvious how this would be assured without a high degree of interpretability, as novel attacks could be conceived of for which the system does not have a track record. However, the false data injection detection system that Niu, Li, Sun and Tomsovic (2019) proposed might be able to solve this issue as it can detect generic false data.

As pointed out in the interviews, power grids involve a lot of dangerous equipment and mistakes could easily have fatal consequences for maintenance workers or other people working near high voltage cables that are supposed to be turned off, which makes the risks involved great and with it the need to trust the system.

An idea that came up in many of the interviews was that the system could gain trust through testing, although even in this case interpretability would allow that trust to be gained quicker. However, as brought up in the section about debugging, testing and validation, it is very difficult to test deep learning systems. It seems that trust will always be an important issue should a deep learning system directly operate the grid. The testing and debugging tool proposed by Sun, Huang, Kroening, Sharp, Hill and Ashmore (2019) might make this possible. But it is unlikely that the power grid business, which, as was pointed out in both the interview with Vattenfall Eldistribution and Energimyndigheten, is quite conservative (with good reason), would accept a fully automated grid operation solution even if there were fool-proof tests. Furthermore, the fact that errors occur infrequently according to interviewees from Mälarenergi, would mean that there are not a lot of opportunities for a grid control system to prove itself.

In the interview with Energimyndigheten, it was stated that for the purpose of long-term forecasts, trust is not particularly important as decisions related to these forecasts would be made by humans who can look at the result critically. However, it is hard to see how the forecasts would be of use if there is not some degree of trust or informativeness. Furthermore, long-term forecasts are often on the scale of decades, which would mean that it takes a very long time before you can see that the results are consistently accurate. Besides, while it is outside the scope of this study, it is worth noting that the potential for deep learning to provide accurate predictions on such a long-term scale is questionable, in part because it would depend on things like population growth in different areas and new technology, which seems difficult to predict merely by using data.

### *5.1.2 Causality*

This issue was difficult to explain to the interviewees and most had trouble answering it since it is technical and complex. The interviewee from Ellevio nevertheless stated that it could be useful for some electricity-quality problems where a large number of factors influence the end result. In general, however, it seems like it would be beneficial, but not of particular importance.

### *5.1.3 Transferability*

All interviewees thought that transferability would be useful, and the interviewee from Ellevio mentioned transfer between network providers as a possible application. One of the interviewees from Energimyndigheten stated that it was not as important as trust. This is in agreement with our own impression that trustworthiness is always important and vital for critical applications whereas the ability to solve other similar problems can be seen as an extra feature which is not always useful. There could be issues where a modified grid would require knowledge transfer, and it seems likely that the system might need human help for the transition, although it would of course be preferable if the system could do it in a fully automated manner, in which case the transfer wouldn’t require interpretability other than for keeping trust.

### *5.1.4 Informativeness*

Informativeness, the ability to understand the reasons behind a decision by a deep learning system, was deemed important by our interviewees but as with trust, it was not deemed necessary. Or at least that is one interpretation of our result. In one instance, the interviewee from Ellevio states that they can conceive of them in the future accepting decisions that they do not understand the reasons for. However, later in the interview, they say that understanding the reasons behind decisions is important during accident investigations. Maybe what they meant was that for the near future, decisions need to be understood but in the long-term, AI systems may become trustworthy enough that there is no need to understand their decisions. Although one could claim that in the long-term we will have developed techniques to understand the decisions AI systems make much better. Suffice it to say that it is unclear whether opaque complexity or understandability will win out in the long-term. The need for investigation was also brought up by Energimyndigheten. They also mentioned that recommendations and forecasts made by a system might be helpful even if it cannot be verified if it is used in conjunction with other information.

### 5.1.5 Fair and Ethical Decision-Making

Although there is some relevance, fair and ethical decision-making is a bigger issue when it comes to systems that are more focused on individuals as there is more data to draw conclusions from and potentially discriminate based on. There are also more decisions that could be discriminatory. While forecasting and grid operation systems could possibly draw conclusions based on power usage to make discriminatory or unethical decisions, the potential should be relatively low when compared to other deep learning systems. There does, however, appear to be some potential problems when it comes to deciding where to improve the grid and selecting for more lucrative customers and while this is quite speculative, it is still important to be aware of.

When it comes to false data injection though, it is not hard to see unfair or unethical decisions accidentally being made. A false data injection detection system could start doing unfair profiling without the owners' or developers' intent or knowledge. For example, it could start predicting false data based on things like location. This could have negative consequences for those alleged to have attacked the system depending on how the grid operator chooses to follow up a suspected attack.

### 5.1.6 Accessability

Accessibility, while not vital, is still important for multiple reasons. One is that it is helpful for gaining trust among parts of the company that are not part of creating the system. This is more suitably done by simple post-hoc explanations than transparency, as the post-hoc explanations can be adapted to be easier to understand while the systems themselves will always be complex. In the interview with Vattenfall Eldistribution this idea of using post-hoc explanations was brought up. Although the interviewee, not knowing the theory around interpretability, obviously did not use that phrase what was described corresponds to the concept of post-hoc explanation. Saliency maps in particular fits well with what was said. It makes sense that saliency maps would be useful for achieving accessibility as they are quite simple to understand.

### 5.1.7 Interactivity

It was not clear from Arrieta et al. (2020) how interactivity relates to black-box issues and interpretability since a system can be made easy to interact with to, for instance enter data, whether or not it is a black-box and it was in this unrelated sense that our interviewees answered the question.

### 5.1.8 Privacy Awareness

As with fair and ethical decision-making, the relevance of privacy awareness is not immediately obvious as the system does not deal with a lot of personal data to draw from. However, the interviewee from Ellevio pointed out that with power use data you can find out a lot of personal information. While it is hard to see any reason why existing systems would have an interest in violating privacy, as personal information is typically only used for billing, it is important to be aware of privacy for future systems with wider functionality. However, it should not be a critical obstacle, but rather something that it will take some extra work to consider and perhaps limit systems to avoid.

## 5.2 Implications for Smart Grid Applications

Several issues arise from the black-box nature of deep learning systems. One issue that was brought up by several of our interviewees was that building trust takes longer if the system can only be verified with ample testing and through experience, since it is close to impossible to formally prove features of the system (Odena and Goodfellow, 2018) or understand its inner workings. The opacity of the system is also a disadvantage when using it to make decisions since operators currently want to know why something should be done. Even if the reasons behind decisions are not needed at the time of the decision, it could be a problem if these reasons are not available afterwards for investigations. However, low trust and lack of informativeness is less of a hindrance for applications with lower risk, for instance analysis of consumption behavior and forecasts. This is in contrast to operation-critical and maintenance systems. The applications we found in the literature, namely forecasting, demand response and false data injection detection, are not operation-critical and should thus not be affected as much from the issues with the inability to interpret the system.

While the post-hoc explanation methods brought up in subsection about categories of interpretability could alleviate some of the issues with interpretability, many of them seem quite theoretical and they all provide a limited explanation. For these reasons it seems unwise to assume that the issues will be fully solved through post-hoc explanations, and the issues thus have to be accounted for when using deep learning for smart grids.

# 6 Conclusion

In general, the black-box nature of deep learning poses a challenge for applying it to smart grids in Sweden. Due to the fact that the Swedish power grid is both very critical infrastructure and can pose a danger to maintenance workers, it will not be feasible for deep learning to directly control the power grid for the foreseeable future. This is because there is a need for the utmost trust that nothing will go wrong and this is not currently attainable with deep learning due to its black-box nature. Its black-box nature arises from its complexity which makes it next to impossible to verify that it will work as intended and not introduce new errors.

There are, however, cases which were suggested during our interviews where deep learning could be useful. For example, the black-box nature is not a huge obstacle when it comes to giving a human operator recommendations on how to control the grid. Similar cases like creating long-term forecasts for planning maintenance and upgrades are also possible. However, in all of these cases a complete lack of interpretability would seemingly make the recommendation or forecasts meaningless as there would be no reason to trust it and no knowledge to be gained. One way to solve this problem would be through achieving informativeness, which would mean that the human decision-maker could take systems reasoning into account in their decision making. Failing that, the system will need to gain trust by proving itself over time, the process of gaining this trust could be expedited with interpretability.

Beyond the application mentioned in the interviews, the applications we found by surveying the literature are not substantially hindered by a lack of interpretability. Trust, which it has become apparent is the most important issue in this area, is not of critical importance to either application as they simply need to generally perform well and not perfectly. As with forecasting and giving recommendations, these applications need a degree of either trust or informativeness for the results to be worth taking into account, but this trust can also be built on a track record. Fair and ethical decision making could be an important issue primarily for false data injection detection, as it affects individuals.

The difficulty to interpret deep learning systems limits their suitability for smart grids by ruling out certain applications like direct control of the grid, mostly because of a lack of trust and informativeness. However, a lack of interpretability does not rule out applications like aggregated load forecasting, long-term forecasting, demand response, false data injection detection and recommendations to operators, although in many cases it does restrict them. Some of these applications, in particular those identified by the interviewees rather than in previous research, may prove challenging to implement for other reasons. Nevertheless, deep learning has potential to benefit smart grids in Sweden despite its black-box nature.

## 6.1 Further Research

In the process of writing this study, we found a number of questions that would be interesting for further research: What kind of private information could be inferred from peoples'

electricity use data? And what type of tasks could lead a deep learning system to make these inferences without its developers' intent or knowledge? How far into the future could deep learning systems reliably predict electricity use before questions which can't reasonably be answered purely with data (such as where people live and what technologies they use) become too important? Research into the possibility for post-hoc explanations to alleviate the identified problems would also be interesting.

# 7 Appendix

## 7.1 Interview Questions (in Swedish)

| Tema | Frågor |
|---|---|
| Bakgrund | • Vad är din roll på organisationen?<br>• Hur länge har du jobbat där?<br>• Hur insatt är du i deep learning eller annan maskininlärning?<br>• Använder ni på företaget för tillfället deep learning för smarta elnät?<br>• Har du erfarenhet av att implementera deep learning?<br>  o isåfall för vilket/vilka användningsområden?<br>• Vet du om deep learning används inom smart grids någonstans idag?<br>• Kan du se någon lämplig användning av deep learning för smarta elnät? |
| Möjlighet för intervjupersonen att ta upp utmaningar | • Om du har erfarenhet med deep learning i smarta elnät: vilka utmaningar har du haft?<br>Om inte: Kan du förutse några utmaningar som skulle kunna dyka upp vid själva implementeringen av Deep Learning för smarta elnät?<br>• Ett stort problem med deep learning är att det är väldigt svårt att förstå hur ett system kommer fram till sina slutsaster, det är vad som kallas en black-box, tror du att det kan vara ett problem inom smarta elnät?<br>  o Vilka problem tror du kan uppstå om smarta elnät förlitar sig på black-box-teknik? |
| Utmaningar från literaturen | • I litteraturen har vi upptäckt 8 teoretiska fördelar med begripliga AI-system, skulle du kunna beskriva hur relevant du tror att varje fördel är för er smarta elnätsverksamhet?<br>  o Tillit, alltså ifall man kan lita på att systemet beter sig som förväntat och framförallt att det inte introducerar nya fel som inte hade uppstått med den tidigare metoden<br>  o Kausalitet, ifall man kan se kausaliteten i en korrelation, till exempel om en modell hittar en korrelation mellan att jag tappar min penna och att det hörs ett ljud, kan man då avgöra ifall pennan gör ett ljud när den landar eller om jag tappar pennan på grund av att jag hör ett ljud? |

| | |
|---|---|
| | ○ Överförbarhet, att man kan använda lärdomar från att ha löst ett problem för att lösa liknande problem, till exempel att man kan använda lärdomar från ett system som kan känna igen personbilar för att känna igen lastbilar<br>○ Informativitet (informativeness), att man får information om hur ett beslut har fattats så att man bland annat kan granska om det är korrekt. T. ex. om ett system granskar en röntgenbild så vill läkaren inte bara se om något ben är brutet, utan även vad som får en modellen att påstå att ett ben är brutet. Dels för att verifiera att det är korrekt, och dels för att veta t. ex. hur allvarligt det är.<br>○ Rättvist och etiskt beslutsfattande, fattar modellen beslut på ett rättvist och etiskt sätt? Oftast kan biaser smyga sig in i deep learning-modeller, både i designen och genom att den tränas på data som är färgad av bias. Det kan till exempel leda till att minoriteter som diskrimineras på arbetsmarknaden också får svårare att få lån för att modellen ser att denna minoritet har lägre snittlön. Deep learning-modeller kan också hitta oetiska kopplingar, till exempel att det är lättare att sälja till någon som är beroende av produkten.<br>○ Tillgänglighet, dvs att andra än experter kan förstå modellen och dennes beslut<br>○ Interaktivitet, att kunder kan interagera med systemet lättare<br>○ Integritetsmedvetenhet (privacy awareness), att man vet ifall modellen kränker personers integritet, t. ex. om den lyckas hitta mönster som kränker privatpersoners integritet värre än originaldatan |
| Andra tankar | • Finns det något annat vi inte tagit upp som är relevant för det här ämnet? |

## 7.2 Interview Questions (in English)

| Theme | Questions |
|---|---|
| Background | • What is your role in the organization?<br>• How long have you worked there?<br>• How knowledgeable are you about deep learning?<br>• Does your company use deep learning for smart grids? [not asked to Energimyndigheten]<br>• Do you have any experience in implementing deep learning?<br>  o If so, in which area(s) of use?<br>• Do you know if deep learning is used in smart grids anywhere today?<br>• Can you see any suitable use for deep learning in smart grids? |
| Chance for interviewee to bring up challenges | • If you have experience with deep learning in smart grids: which challenges have you faced?<br>If not: Can you predict any challenges that would arise at the implementation of Deep Learning for smart grids?<br>• A large problem with deep learning is that it's very difficult to understand how a system arrives at its conclusions, it's what's called a black-box, do you think that this could be a problem in smart grids?<br>  o Which problems do you think would arise if smart grids relied on black-box technology? |
| Challenges from the literature | • In the literature we have discovered 8 theoretic advantages with comprehensible AI-systems, could you describe how relevant you think each advantage is for your smart grid business?<br>  o Trust, that is if you can trust that the system behaves as expected and especially that it does not introduce new errors that wouldn't have arisen with the previous method.<br>  o Causality, if you can see the causality in a correlation, for example, if a model finds a correlation between me dropping my pen and sound being heard, can you then determine whether the pen makes a sound when it lands or if I drop the pen because I hear a sound?<br>  o Transferability, that you can use learning from having solved a problem to solve similar problems, for example that you can use learning from a system that can recognize private cars to recognize trucks.<br>  o Informativeness, that you can gain information about how a decision has been made so that you can among other things determine if it's correct. For example, if a system reviews an x-ray image, the doctor does not |

| | |
|---|---|
| | just want to know if a bone is broken, but also what makes the model say that a bone is broken. Partly to verify that it's correct and partly to know for example how serious it is.<br>○ Fair and ethical decision-making, does the model make decisions in a fair and ethical way? Often biases can sneak their way into deep learning models, both in the design and because it's trained on data that's coloured by bias. It could for example lead to minorities who are discriminated against on the job market also having more difficulty getting loans because the model sees that this minority has a lower average wage. Deep learning-models can also find unethical connections, for example that it's easier to sell something to someone who is addicted to the product.<br>○ Accessibility, that is to say that others than experts can understand the model and its decisions.<br>○ Interactivity, that customers can interact with the system easier.<br>○ Privacy/integrity awareness, that you can tell if the model violates persons' integrity, for example if it manages to find patterns that violate private persons' integrity worse than the original data. |
| Other thoughts | • Is there anything else we haven't brought up that is relevant for this subject? |

## 7.3 Transcription of Interview 1

J = Jesper Lundberg
A = Alexander Lundborg
P1 = Interviewee from Vattenfall Eldistribution

**1. J: då börjar vi med lite bakgrund, vad är din roll på vattenfall?**

**2. P1:** Eh, först så jobbar jag på vattenfall eldistrubution, alltså nätbolaget och det är ju [inaudible] skäl då inte helt i alla avseenden samma som vattenfall men på vattenfall eldistribution är jag chef för en grupp som heter digital hub som ansvarar för digitalisering och innovation.

**3. J: Ok, hur länge har du jobbat där?**

**4. P1:** ja på vattenfall har jag jobbat sedan 2003 men just det här specifika uppdraget, ett år.

**5. J: Är du något insatt i deep learning eller annan maskininlärning?**

**6. P1:** Nä, jag är ingen expert på deep learning och machine learning, det är jag inte. Däremot gör vi projekt och innovation som tillämpar det här så i det avseendet så känner jag till det men i detaljer om, vad man ska säga, någon akademisk natur i det här, det kan jag inte, men effekterna förstår jag.

**7. J: Så använder ni på företaget, eller du nämnde ju det att ni använder lite deep learning för tillfället för smarta elnät?**

**8. P1:** ja alltså vi gör ju kontinuerligt... alltså det stora fokuset de senaste åren är egentligen att använda den information man har på mer och mer sätt och allt från trivial data-analys till mer avancerade algoritmer och machine learning med mera, där man försöker ta nästa steg. Allt är ju inte, det är ju fortfarande, iallafall i termer av elnätsverksamhet tämligen ny verksamhet så att det är ju inte allt som man kommer på allt som är bra är ju inte alla gånger utrullat i bred användning, men vi har ett antal fall där vi analyserar data, ja avancerad data-analys helt enkelt för att få ut värden i olika avseenden.

**9. A: Är det just deep learning som ni använder eller är det någon annan typ av avancerad data-analys?**

**10. P1:** det är ju, i den mån jag förstår det då så är det ju machine learning som är att, och mycket kopplat till mönsterigenkänning.

**11. A: Ok, och vad använder ni det till? Är det för smarta elnät eller?**

**12. P1:** ja, det är ju egentligen två, mycket handlar om att analysera där vi har stora datamängder och typiskt så är det inom områden också där man hamnar väldigt nära kunden alltså antingen den data som samlas in på de smarta mätarna eller både detaljerad och aggregerad förbrukningsdata för att uppnå prognoser för framtiden, förbrukningsprognoser på olika områden. Ni känner kanske till att det finns på vissa ställen kapacitetsutmaningar i elnäten. Skåne är ju ett sånt område, Uppsala är ett annat sånt område och Stockholm ligger inte långt därefter bara för att nämna några då, där bygger väldigt mycket på dem aktiviteter

man gör... förmåga att kunna säkerställa rätt prognoser för kommande behov och det är både i det korta perspektivet alltså de närmsta timmarna, dagarna men också i ett lite längre perspektiv, dvs år och till och med decennium då. Och man kan ju säga såhär att ju längre bort i tiden man kommer desto mindre värde kan olika former av machine learning och deep learning fungera därför att det bygger på så många okända parametrar som inte riktigt speglas i tidigare mönster men däremot när man tittar på prognoser på timmar och dagar då finns det med rätt intrimmade metoder så ser man en tydlig förbättring över tid i de prognoser man får när algoritmerna får jobba.

**13. J: Kan du säga några utmaningar ni har haft med deep learning eller maskininlärning, för smarta elnät just?**

**14. P1:** Alltså jag tror ju en utmaning är att det är ett nytt begrepp och det finns ju ett, och det är ju inte bara när det är machine learning utan är det i någonting som kan betecknas som konservativ branch så är det ju oftast att få en acceptans för att den information man får ut, och det behöver inte bara [vara] av machine learning eller deep learning-metoder... att man kan lite på dem. Så att det är ju liksom en förtroendeutmaning som är det som kanske tar längst tid mer än det rent tekniska.

**15. J: det va lite det vi tänkte komma in på nu faktiskt, för att som du kanske vet så är det svårt att förstå varför deep learning får sina resultat, det är alltså en black-box. Har du koll på det?**

**16. P1:** Nä alltså jag kan ju inte detaljerna i det här så det kan jag inte svara på men det är ju alltid så här, är det någonting som man inte absolut kan validera, att det här är rätt och man förstår det, så finns det en större skepsis och det är svårare att få en acceptans för det men det här har inte bara med deep learning och machine learning att göra utan jag skulle kunna ta vilken annan tjänst som helst, det kan vara något väldigt väldigt trivialt. Om jag lägger ut det till ett externt bolag till exempel och köper det som en tjänst men jag får inte berätta va de gör med informationen även om de bara tar A plus B så har man samma utmaning. Bara för att ta ett helt annat exempel, vi har ju i snart ett decennium haft vårt Scada-system, alltså det som styr elnätet rent operativt, haft förmågan att det systemet i händelse av fel kunna isolera det felet helt själv, alltså dvs koppla bort, öppna brytare, frånskiljare, på ett sätt så att så få kunder som möjligt blir påverkade av det här. Och systemet har klarat av att göra det här i ett decennium men det är ju inte det som är det viktiga, det viktigare är att det är operatörerna som har elsäkerhetsansvaret som måste lita på att systemet kopplar rätt och det finns områden där det är sånt att man fortfarande inte litar på systemet. Det har ingeting med deep learning och machine learning att göra utan, det är en trevlig funktion, utan det har med förtroende att göra där det finns ett ansvar för säkerhet men det kan finnas finansiellt ansvar osv också som gör att litar man inte på vad man får ut så kommer man inte använda det.

**17. J: det jag menar är att det är ofrånkomligt om man använder deep learning, det går inte att se varför det händer (varför ett beslut fattas)**

**18. P1:** så är det ju, och där är den tråkiga verkligheten eller erfarenheten hittintills skulle jag vilja säga är att tid löser det här men det behövs oftast rätt mycket tid, inte för att algoritmen tar tid utan för att förtroendet för den svarta lådan som du säger ska uppstå och det är klart att hitta en genväg runt det där det finns säkert saker man kan göra. En av grejerna som jag tror är rätt viktig är ju att man ska förstå hur lätt det är att störa de här algoritmerna baserat på indatat. En dags felaktiga värden, hur stort genomslag får det på en prognos. Någon sån här

robusthet, jag vet inte det kanske inte ens är en term i det ni håller på med men jag skulle vilja formulera liksom det att algoritmerna är robusta, att ett fåtal felaktiga värden inte vänder uppochner på världen.

**19. J: Jo men robusthet vet jag att jag stött på i litteraturen och det finns ju vissa metoder för att göra det lättare att förstå tex se vad som är viktigt så att man kan se om det här kommer bli fel men det verkar vara ganska teoretiskt fortfarande. Vi har hittat åtta fördelar i literturen som kan uppnås om man skulle kunna göra det lättare att förstå deep learning. Kan vi fråga dig hur viktig du tror att varje sådan fördel är just för smarta elnät?**

**20. P1:** Ja, absolut

**21. J: Vi har redan varit inne på tillit, alltså att man kan lita på att systemet gör rätt och att det inte introducerar nya fel är framförallt viktigt. Är det något du vill lägga till där eller ska vi gå till nästa?**

**22. P1:** Du kan ta nästa tror jag

**23. J: Sedan så är det kausalitet, om du förstår hur systemet funkar kan du förstå kausala samband, så att man inte bara ser en korrelation utan att se vad som orsakar vad. Tror du det hade varit viktigt för smarta elnät?**
**24. P1:** Ja alltså det tror jag och framförallt kanske, det blir ett orsaks-samband här, är det någon som jag känner och litar på alltså på bolaget som har skaffat sig den förståelsen och förstår det här i detalj så kan jag nog vara benägen att lita mer på det. Så någonstans så kan man ju bygga förtroende genom att någon man litar på känner sig trygg med det och då kan jag också lita på det.

**25. J: Sedan har vi överförbarhet, som är att man kan använda lärdom från att ha löst ett problem till att lösa något likande problem. Kan du kunna komma på om det skulle vara relevant?**

**26. P1:** Det tror jag säkert, jag skulle vilja säga att vi kanske inte befinner oss där än utan det är fortfarande kanske i så stor linda att använda den här typen av metoder men jag tror absolut det kan ha påverkan, att kan vi säga att nu gör vi det här som bara är en liten utökning på vad vi gjorde tidigare. Dessvärre har vi inga sådana fall idag så det blir en ren spekulation för mig men jag tror absolut att det kan ha en påverkan på vad vi gör framöver.

**27. J: Och sen så hade vi informativet vilket handlar om att kunna granska beslutet, ett exempel är att en läkare vill inte bara ha en AI som säger att patienten har cancer utan han vill se varför systemet tror det så han kan se om systemet har fattat beslutet på rätt sätt.**

**28. P1:** Ja alltså någon form av granskning innan det exekveras eller användes det tror jag är en del av resan så att säga och jag kan ju dra parallellen till det här att isolera fel i nätet är ju ett, vägen framåt som man gör är att... man går ju inte från en manuell process till full automatik utan man går till någon form av halvautomatik. Liksom, berätta för mig vad du skulle ha gjort och så säger jag okej till dig, det blir lite samma sak som jag uppfattade det där med läkaren, att okej ge mig rådet men det är jag som bestämmer om vi ska följa det.

**29. J: Tror du förresten att det här är... att eftersom att det handlar om ganska viktig infrastruktur, att det är viktigare just med elnät att man har den insynen.**

**30. P1:** Ja, asså det, asså, jag tror ju liksom, man får analysera vad är risken eller vad är den värsta konsekvensen det kan bli. Och jag menar är de värsta konsekvenserna att någon dör, ja då är det ju klart att då är man ju mer försiktig än om värsta konsekvenserna är att man förlorar lite pengar, coh någonstans mittemellan är att man förlorar mycket pengar. Så det är ju kalrt att det finns en risk... en skala med olika risker man får beakta. Men kommer man ner till säkerhets, så är man ju, så finns det ju en väldigt låg benägenhet att acceptera risker. Och är man med smarta elnät nu, för det är ju ett begrepp som vi egentligen inte har pratat om vad är det egentligen, men om man isolerar det till rent att styra i nätet så finns det ju en personsäkerhetsaspekt i det här som det finns en väldigt låg vilja att utmana eller riskera.

**31. J: Ja, nästa faktor här: "rättvist och etiskt beslutsfattande", det handlar ju om att beslut riktade till personer så kan det liksom komma in biaser från, antingen träningsdatan, eller hur man har skapat algoritmen, ehh, och beslutsfattande kan också, det kan också handla om att man liksom gör oetiska val, till exempel att man hittar så här att, nån sån här marknadsföringsbot kanske hittar att nån är beroende av nåt och tycker att jamen då ska vi marknadsföra det till den personen. Tror du att det, så om man har insyn i systemet så kan man ju se att det fattar sådana beslut, tror du det är något som kan vara en faktor här?**

**32. P1:** Tveksamheter för elnätsverksamheten tror jag inte att det är särskilt överhängande, eftersom det finns, det är klart med tillgången till, mät- förbrukningsdata från kunder, till exempel så är det klart att man kan dra slutsatser utifrån dem. Men det är ju också, det är ju också den anledningen varför man har det regelverk med unbundling och att elnätsbolagen är separerade från övriga aktörer, för att egentligen motverka det här, så för elnätsbolagen, givet att man inte gör ett brott och bryter mot det regelverk som finns så det ska, det gör man inte och det ska man inte göra så, så, så finns det ju inte, den informationens enda värde är ju att förbättra den tjänsten kunden redan har, vilket gagnar kunden i slutändan. Det finns ju inte en ny produkt man kan sälja till, till alltså det finns ju inte fler produkter, utan produkten är ju elleverans och ingenting annat.

**34. J: Ja, nä som sagt vi ville ju inte liksom göra antaganden om vad som är relevant, utan vi valde att fråga dig om allt, eller liksom.**

**35. J: Ehm. Nu kommer jag inte ihåg hur. Ah, juste, tillgänglighet innebär att folk som inte är särskilt insatta eller inte är experter kan förstå eh, ja, liksom hur beslut har fattats, så att ofta om man är expert kan man förstå lite grann hur liksom hur deep learning-algoritmen har fungerat och hur beslutet har fattats. Ehm, men så att om man kan lyckas göra det lätt att förstå kan det gå för liksom folk som inte är så insatta att också förstå. Tror du det är viktigt?**

**36. P1:** Asså jag tror att i alla sådana här fall, när, för att det finns ju många av de datasystem som vi använder idag där alla inte kan förstå, men jag tror att man måste anstränga sig att göra det, att kanske inte förklara algoritmerna, utan mer förklara att utifrån den här datan så får vi de här slutsatserna. Och kan man förklara det på ett tydligt och också kan beskriva hur känslig den här algoritmen är mot felaktiga data, felaktiga indata. så tror jag man kan bygga förtroende, men återigen, som jag sade tidigare så tror jag ju att, man kan nog hjälpa tiden på traven, alltså göra det snabbare genom att göra bra förklaringar, men jag tror generellt sett att i

områden där det finns en hög risk så tar det här tid, vi pratar år innan man kan bygga upp stort nog förtroende för man ska lita på det fullt ut. Sen kan man ju ha det här som ett stöd, i den rådgivande funktionen som vi nyss pratade om. Det tror jag kan finnas, det tror jag att många inte har något stort problem med. Men det här att man litar på det fullt ut, det vill säga att man har någon form av automatik runt användningen av vad man får ut av machine learning, det tar tid innan man…

**37. A: Kan jag följa upp på en fråga, hur uppnår man förtroende för de här, nya metoder för deep learning, är det att man... hur validerar man det eller hur får man förtroende för dem.**

**38. P1:** Ja, asså återigen, det viktigaste tror jag är att liksom att man... att de som ska få förtroende får se och känna på det här så att säga, under en längre tid, sen kan man säkert snabba upp det där med att förklara saker runt det, men, ehh, men, det här handlar väldigt mycket om att de som ska använda det här ska lite på vad man får ut, och då tror jag väldigt mycket på trial and error, eller förhoppningsvis inte error, men att man får iallafall testa och känna, och över tid får man ut att, ja men det här är ju rätt och det litar jag på och så vidare. Jag tycker det är jättebra om ni kommer på någon genväg och man kan säkert snabba på det på olika sätt genom att förklara och stödja de som behöver lita på det här, men i grund och botten tror jag det är en fråga om att man måste ge det tid.

**39. A: Och då är det att se att till exempel prognoserna som systemet skapar är trovärdiga eller tillförlitliga eller korrekta?**

**40. P1:** Ja, precis, det är ju, för just den tillämpningen, ja. Asså att man kan se att det här stämde ju, att jag fick rätt råd och... och sen över tid så, ja men det här är rätt. Jag menar vi kan ju ta, de prognoser vi har haft, jag menar man kan ju se att det blir ju aldrig exakt rätt, för jag menar det finns ju en viss stokastisk, verkligheten är ju lite stokastisk av sig själv så att säga. Men liksom, själva algoritmerna tar ju tid på sig. Nu har jag inte den exakta procentsiffran, men det var ju, det var ju rätt avsevärda förbättringar över ett kvartal, där det var rätt dåligt dag 1 och dag 90 var det väsentligen bättre, så även där behöver du ju tid för att algoritmer ska känna att de förbättrar sig, och det blir bra slutsatser, då får vi se här, jag menar har vi kört dem i flera år kanske det blir ännu bättre.

**41. J: Det är bra, ska vi gå vidare då?**

**42. A:** Ja

**43. J: Näst sista faktorn var interaktivitet, vilket innebär att det är lättare för kunder att interagera med systemet, ehh. Jag vet inte om det är något som skulle va relevant här. Att kunder ska…**

**44. P1:** Ne jag... Jag förstår inte riktigt, jag förstår inte riktigt

**45. J: Nä, vad, lite vagt, jag vet inte är det någonstans där liksom kunder kommunicerar med era system**

**46. P1:** Ja, asså vi har ju, jag menar det är ju kalrt att vi har sett att kunder kommer åt sin förbrukningsinformation och räkningsinformation och så vidare. Och det är klart att vi har också sätt för industrikunder att lägga sina prognoser som hjälper oss. Så det är klart att det

finns en, det är klart att om du vill. Jag är inte helt säker på din fråga, men liksom det är klart att kan våra kunder ge mer information än annars var tillgängligt, så kan vi göra till exempel våra prognoser eller slutsatser kan ju bli bättre där. Det bygger ju på en frivillighet, så det är klart att med data i andra, med lite andra dimensioner kan förbättra de här svarta lådorna. Jag menar till exempel är det bättre att en kund lägger in sin förväntade förbrukningsprognos, en industris förväntade hur de kör, och de ger det till oss, än att vi ska försöka på deras tidigare förbrukning hur de tänker köra. Jag menar så är det, men samtidigt kan vi inte förvänta oss att alla gör det. Så jag vet inte riktigt om det var det du åsyftade, men det är klart att man kan hjälpa med de svarta lådorna om man får mer exakt data.

**47. J: Ja, men det låter bra. Ja, det var ungefär det. Och sedan var sista faktorn integritetmedvetenhet, att om systemet hanterar personlig data så kan det liksom råka kränka folks integritet, och ja om man ser hur det fungerar kan man se till att det inte gör det, är det något som du tror kan vara en faktor här?**

**48. P1:** Ja alltså det är ju oerhört viktigt att man följer det regelverk som finns, GDPR med flera, så det är ju det basala, för det är där det ska byggas förtroende för att vi inte använder den information som vi har om kunderna på ett otillbörligt sätt. Det behöver ju inte betyda att kunderna litar på det så att säga, men det är ju tanken med det regelverket att skapa den gemensamma grunden, att man ska veta att det sker på ett schysst sätt.

**49. J: Men hanterar ni, för att liksom det som kan vara problemet är ju till exempel om vi skulle ha något helt annat system som kanske jag vet inte, det kan liksom kanske en marknadsföringsbot som baserat på sociala medier hittar liksom partitillhörighet som de använder för att marknadsföra något så har de ju helt plötsligt hittat nån liksom, extra skyddad data enligt GDPR. Jag vet inte, hanterar ni liksom mycket persondata, för att det skulle kunna hända sådana saker, tror du?**

**50. P1:** Inte som du, jag menar det är klart att vi har persondata. Jag menar vi vet ju vad kunden heter och var de bor någonstans och vi vet ju den förbrukning som sker. Däremot så används ju inte den på något annat sätt än att, som identifierare, som enda sättet att skicka fakturan. Sen så sker ju en aggregering av den informationen, men då försvinner ju den personkänsliga biten.

**51. J: Okej, men då tror jag vi var klara där.**

**52. A:** Jag tror bara vi har en sista fråga bara som jag såg här. Om vi har något annat som är relevant för ämnet som du vill ta upp, om du har någonting.

**53. P1:** Nej, jag tror ju väldigt mycket att det här är ju, olika former av dataanalys, tror ju jag kommer, det är ju inte bara machine learning, deep learning som kommer vara, utan alla typer av dataanalys kommer ju bara öka och jag tror ju inte att det sättet och de slutsatser vi idag tror oss kunna använda det här till kommer vara det som i förlängningen kommer vara begränsade av, utan jag menar det är en stor värld som öppnas för det kommer ju ske utveckling under många år inom det här området tror jag. Vi är ju egentligen bara i början på en resa skulle jag vilja säga. Så att man ska inte, jag menar man får inte dra, det är inte ändstationen vi befinner oss nu på. Så att möjligheterna är ju fortfarande många. Och det sker väldigt mycket arbete för att kunna göra mer och mer avancerad dataanalys. Och det är inte bara inom elnätsbranchen, utan det är hela vårt samhälle som ni säkert känner till. Nej, annars

har jag inget mer att tillföra, så lycka till, med ert arbete. När ni är klara får ni gärna skicka mig en kopia av det. Bra, men ja, ha det bra!

**54. A: Tack så mycket.**

**55. P1:** Hej.

## 7.4 Transcription of Interview 2

J = Jesper Lundberg
A = Alexander Lundborg
P2 = Interviewee from Ellevio

**1. J: Vad är din roll på organisationen?**

**2. P2:** Jag är utvecklingsingenjör på en avdelning som heter lokalnät Stockholm. Jag arbetar framförallt med teknikfrågor rörande lokalnätet.

**3. J: Hur länge har du jobbat där?**

**4. P2:** Jag började 2011

**5. J: Du sa att du inte var insatt i deep learning eller annan maskininlärning va?**

**6. P2:** Korrekt

**7. J: Vet du om ni på företaget använder deep learning för smarta elnät?**

**8. P2:** Eh, alltså för smarta elnät vet jag inte om vi gör det eller då tror jag inte vi kan säga att vi gör det men däremot gör vi det ju för att ta fram prognoser. Ni blev tipsade om [person] också, jag vet inte om ni kommer få till någon intervjun med henne eller om den informationen nådde fram till er.

**9. A: Ja, jag har skickat ut mail till henne men jag har inte fått svar**

**10. P2:** Ok då kan det vara så att hon är för upptagen, men hon har det senaste året, två åren, tagit fram maskininlärningsalgoritmer för att kunna göra prognoser för effekten.

**11. A: Kan jag fråga är det hur mycket effekt som efterfrågas eller som produceras?**

**12. P2:** Jag tror... det är efterfrågad effekt som hon prognostiserar

**13. A: Ja, ok**

**14. J: Vet du om det är lång eller kort sikt?**

**15. P2:** Både kort sikt och lång sikt

**16. J: Ok, kan du med det du vet se någon lämplig användning av deep learning i smarta elnät?**

**17. P2:** För det första tror jag att jag behöver förstå skilladen mellan maskininlärning och deep learning, är det någon... det är förmodligen någon skillnad eller?

**18. J: Ja, deep learning är en sorts maskininlärning. Maskininlärning kan vara, det är lite mer odefinierat. Det kan vara liksom... linjär regression kan ses som en sort maskininlärning. Deep learning är just när man har ett neural network som är bättre på att se mönster.**

**19. P2:** Nä men exakt och då tror jag att det kommer säkert finnas möjliga applikation för maskininlärning och deep learning. Däremot kan man säga att vårat system är ganska primitivt när det gäller digital kommunikation, spara ner och logga data, och även för att manövrera object. Och det är ju det vi jobbar, som jag jobbar mycket med och har jobbat mycket med de senaste åren är ju att bygga

upp ett system för att kunna samla in information så vi kan övervaka elnätet, så att vi kan fjärrmanövrera elnätet och för att sedan automatisera det. Och där ser man ju absolut möjliga, möjligheter för att kunna i framtiden använda maskininlärning.

**20. J: Mm, bra. Nu kommer vi in på det vi nämnde med black box, att det är svårt att förstå varför ett deep learning-system får sina resultat och hur det funkar där inne. Skulle du själv kunna komma på några problem som kommer uppstå? Vi kommer senare fråga om vissa specifika problem som litteraturen tagit upp.**

**21. P2:** Jo men exakt, det låter ju som att det kan bli ett problem om man jämför med dagens processer och hur vi hanterar systemen, så bygger ju allt på att vi ska ju veta exakt, liksom, varför vi gör någonting. Vi skriver ju sånna här driftsedlar för att, när det sen ska manövreras så ska det exakt vara tydligt varför man gör en viss grej och även då driftpersonalen vill ju liksom förstå varför ett beslut tas. Däremot så tror jag att man... jag kan ändå se möjligheter i framtiden att det till exempel presenteras, liksom, förslag för en driftoperatör som de kan acceptera utan att kanske förstå exakt varför. Men de måste förstå, liksom, att slutresultatet blir bra. Till exempel då såhära, vi har ett strömavbrott och idag då kollar vi på lokalnätet så, mellanspännningsnätet, så tas ju, då tittar man och så försöker man liksom hitta den bästa lösningen på problemet och det är ju som ni förstår mycket inparametrar som man kan ha i ett sånt läge. Har man då en automatiserad funktion som liksom nästa steg för oss är, då är ju egentligen, då sker det en automation, egentligen som bara bygger på liksom algoritmer, “om den är stängd, öppna den”, och så försöker man ta det optimala, och då presenteras ett förlag för driftoperatören som säger såhär: "accepterar du den här kopplingen?" nästa steg det skulle ju vara att operatören inte ens behöver acceptera utan det bara sker och i det läget så, nä visst jag tror inte man kommer behöva förstå varför utdatan blev som den blev men man ju, det handlar väl om det här med tillförlitlighet, liksom att du har sett åh fan det blev bra liksom även fast vi inte förstår exakt varför det gjordes på det sättet som det gjordes.

**22. J: Vad bra, när man försöker göra deep learning mer förståbart så är det åtta grejer som man vill uppnå. Så vi tänkte fråga dig hur viktigt du tror att alla dem faktorerna är för smarta elnät. Vi har aktivt valt att inte sålla i dem så vissa kan kännas liksom att det är konstigt för smarta elnät men det är för att vi inte velat göra antaganden**

**23. P2:** Vad var frågeställningen till mig då, huruvida jag tror det...

**24. J: Huruvida du tror det är viktigt för smarta elnät**

**25. P2:** Ok

**26. J: Så vi har ju redan vara inne på tillit, alltså att man kan lita på att systemet beter sig som förväntat och framförallt är det viktigt att det inte introducerar nya fel då, är det något mer då känner att du hade velat säga där eller?**

**27. P2:** Gällande tillförlitlighet?

**28. J: Ja**

**29. P2:** eh, alltså det är ju en väldigt viktig punkt i och med att vi, det som är våran viktigaste grej det är ju personsäkerhet, liksom att antigen att skydden i våra stationer eller i liksom nätet gör det dem ska för att skydda personer och även att när vi gör driftomläggningar eller fjärrmanöver eller likande, så har vi, då kan vi va i ett läge där vi har personal som är ute och arbetar i elnätet så då är det också så här... tillförlitligheten är ju riktigt viktig.

**30. J: En annan faktor är kausalitet, alltså att man kan se kausaliteten i en korrelation. Tror du det kan vara viktigt för er? Alltså om ditt system hittar en korrelation, så kan du ju inte bara på**

**det se vad som är orsak och verkan. Är det något som du tror skulle vara viktigt för er att kunna se?**

**31. P2:** Alltså orsak-verkan kopplingen, menar du?

**32. J: Ja**

**33. P2:** Det tror jag är viktigt absolut för i och med att vi har ett stort system som hänger ihop hela vägen från producent via stamnätet, regionnätet, lokalnätet, ut till slutkund så är det väldigt många faktorer som kan påverka slutresultatet och ser man då tillexempel elkvalitets-problem eller liknande så är det viktigt med orsak-verkan, ja.

**34. J: ok, sen är det överförbarhet som innebär att om ett deep learning-system har lärt sig att lösa ett problem så kan man överföra den kunskapen för att lösa liknande problem. Om man ser hur systemet fungerar så kan man anpassa det för liknande problem.**

**35. P2:** Går det att vara mer konkret, bara så att jag förstår. Skulle det va såhär att vi hittar ett system för att förutsäga fel på våra kablar för en spänningsnivå så skulle vi kunna överföra det till en annan spänningsnivå eller?

**36. J: Ja det skulle det kunna vara, tillexempel. Det exempel vi såg som inte är relaterat til elnät är om man har ett system för att identifiera bilar så kan man använda det för att lära ett system att identifiera lastbilar.**

**37. P2:** Exakt, sen kan jag tänka mig liksom att i våran bransch kommer det ju finnas mycket såna, så att företag kommer kunna överföra lösningar från ett elnätsföretag till ett annat, alltså typ en leverantör av det här skulle ju kunna använda, i och med att vi har, alla systemen påminner ju om varandra även om alla elnätsföretag är unika i sättet man löser vissa grejer men på det stora hela så är det ju samma grej. Så det kommer nog finnas många sådana överföringsmöjligheter.

**38. J: Ja, och sen så har vi informativitet, som innebär att man kan. Att man får information om hur beslut har granskats så att man kan, bland annat har det här fattats på korrekt sätt. Går ju lite tillbaka i tillit, ett exempel som är orelaterat till elnät är liksom en läkare får en bild på en såhär, scanning så vill han inte bara se att den här personen har cancer, utan vill se varför har den här personen cancer så att han kan liksom se hur ska man behandla det, och liksom stämmer det.**

**39. P2:** Det asså, vad kallade du punkten?

**40. J: Informativeness var det på engelska, som vi översatt till informativitet.**

**41. P2:** Aa. Jag tror att den är ju jätteviktig för oss. Allt handlar ju om, vi har ju ofta behov av att gå tillbaka och titta på vad som har hänt, liksom varför blir det ett fel, ett större avbrott? Även om vi inte i stunden hinner att analysera för att vi ska lösa situationen så kommer det sedan att göras typ en haveriutredning där man går igenom och exakt i detalj liksom förklarar, återskapa vad som har hänt och varför. Och då tror jag att den här, att det finns dokumenterat varför vissa beslut gjordes, om det är av en människa eller om det är då av en dator är viktigt.

**42. J: Ja, ok. Sen så har vi rättvist och etiskt beslutsfattande, som framför allt handlar om när man fattar beslut riktat mot en person så kan det liksom läggas in biaser i systemet, eller det kan tränas på data som liksom reflekterar, är liksom orättvis, att ett system kan liksom till exempel diskriminera vissa grupper. Eller så kan det hitta mönster som leder till oetiska beslut, till exempel en marknadsföringsbot kan marknadsföra alkohol till en alkoholist, för att det säljer**

**bra, och det blir ju inte etiskt. Och ser man då hur systemet funkar så kan man liksom hitta att det händer och stoppa det. Är det något du tror kan vara en faktor här.**

**43. P2:** I och med att vi jobbar mycket, det kommer ju vara maskin till maskin eller vad man ska säga. Vi har en dator som får information från vårat system för att sedan ta ett nytt beslut och skicka ut till det här systemet, så har man ju lite svårighet direkt att se den parallellen. Sen skulle man ju kunna tänka sig mer med att det sker ju ständiga upphandlingar av nya komponenter och så man skulle ju kunna om drar det lite längre se att kanske vissa leverantörer skulle kunna favoriseras eller liknande, om man bara spekulerar. Den risk jag ser är då mer att det skulle kunna ge fördel åt vissa företag. Även om den risken känns som liten nu, men det är väl det jag ser.

**44. J: Ja, jag kan ju säga nu efteråt att det var lite på grund av den här som jag kände att jag skulle påpeka att vi inte sållade i förväg, men ja det är bra tänkt på det där med företag, det hade jag faktiskt inte tänkt på innan. Ja och sen så har vi.**

**45. P2:** Asså bara för att förtydliga eller komplettera det där. Typ så att man har en leverantör som levererar den här maskininlärningsdatorn eller deep learning-datorn till oss, och den har samma, det företaget har samma ägare som även äger en större leverantör eller en tillverkare av elnätskomponenter, så skulle man kunna se den risken att man framhäver liksom sina egna tekniska lösningar.

**46. J: Jo, men det är sant. Ja, sen så har vi tillgänglighet som innebär att folk som inte är särskilt insatta i systemet och inte är experter kan förstå det.**

**47. P2:** Exakt, och den är ju viktig, den är ju jätteviktig. Mycket av det vi jobbar med handlar ju om att göra komplexa system lättanvända. I och med att vi som Ellevio vi är ju ett renodlat egentligen, en renodlad beställningsorganisation, vilket gör att vi måste hålla uppe kompetensen för att kunna skriva specifikationer, kunna veta hur vi underhåller vårt system och hur vi driftar vårt system. Men sen jobbar ju vi mot leverantörer som ska leverera till oss, och det måste vara ganska enkla gränssnitt, relativt, och vi jobbar ju med kunder, där ska det ju vara ett hål i väggen, det ska vara extremt enkelt, de ska inte alls behöva tänka på det. Och vi jobbar med våra entreprenörer som är ute i fältet, både när de bygger station, underhåller, eller bygger nätet, elnätet, underhåller elnätet och framför allt då driftar elnätet, till exempel vid ett fel. Och det ska va väldigt enkelt, så att jag tror att ett sådant här system implementerat hos Ellevio skulle vara antingen att vi har liksom den hårda kompetensen hos leverantörer så att vi har ansvariga inom bolaget som jobbar mot dem eller att vi har vissa kärnkompetenser vill vi ju även hålla internt. Så att man skulle kunna tänka sig en grupp som är liksom extremt duktiga även inom Ellevio, men för den vanliga användaren både på Ellevio och hos våra sammarbetspartners måste gränssnitten vara extremt enkla.

**48. J: Ja, ok. Ja. Och sedan så har vi interaktivitet, som innebär att kunder kan interagera med systemet lättare. Ja juste det var ju det du kom in på där, att ja det ska bli lättare att interagera. Och så har vi.**

**49. P2:** Jo men exakt, för där kan man ju tänka sig till exempel då om vi skulle ha ett system, en sån här tjänst som efterfrågats är att, att vi ska vara tydligare eller mer transparenta med var det finns kapacitet i nätet, och det kan till exempel vara en solcellsproducent eller nåt företag som vill bygga upp solcellsanläggningar runtom i Sverige, då vill de, för det spelar inte så stor roll vart de installeras, det är klart, det är bättre ju längre söderut, men de är ganska flexibla vart de sätter de här installationerna, och då skulle de vilja ha ett ganska enkelt gränssnitt för att titta så här: här i det området finns det kapacitet, ok och så kan de börja kolla med markägare där. Och det är ju typiskt en sådan grej där det skulle kunna vara för slutanvändaren då ett ganska enkelt system men bakom det är extremt komplext, och det är därför vi inte har det där idag, för det är så mycket faktorer som spelar in. Det är så mycket data som ska in för att göra de bedömningar, så att vi måste liksom göra det. Eller

idag har vi inget system som kan hantera det utan vi måste ha en människa som tittar på det och tar in alla parametrar.

**50. J: Ja, sen var sista faktorn här integritetsmedvetenhet, vilket innebär att, det är framför allt om man hanterar stora mängder persondata så kan det råka, man kan råka hitta mönster som kränker folks integritet, eller att systemet liksom på något vis inskränker folks integritet. Och då om man vet hur det funkar kan man hitta det och stoppa att det händer.**

**51. P2:** Ja, men det är, exakt. Det handlar ju om att vi, eftersom att elförbrukningen säger extremt mycket om en individs liv så är det ju en av våra delar som gör att vi måste vara extremt konservativa när det gäller till exempel att lämna ut mätdata inom forskningsprojekt och så. För att du kan ju få se exakt när folk går till jobbet, när de kommer hem och om de varit sjuka någon dag. Så att, vi kommer ju gå mot timupplöst data hos slutkunder, idag är det ju frivilligt om du vill ha timdata eller inte, så att vi loggar ju inte timdata överallt , i framtiden kommer det ske och då kommer man ju få extremt bra bild över våra en miljon kunder. Och i USA det kanske ni har läst om där har ju man använt det här för att kunna då hitta marijuana-odlingar genom att helt enkelt titta på data på elmätarna och se där elförbrukningen sticker ut och då åker de dit och så har de hittat några ordentliga lampor som står och lyser på marijuana-odlingarna i garderoberna och sådär.

**52. J: Ok. Ja, det var det vi hade att fråga om egentligen. Är det något mer du känner att du vill ta upp eller så?**

*the conversation continues on topics that are not relevant for the study*

## 7.5 Transcription of Interview 3

J = Jesper Lundberg
A = Alexander Lundborg
P3 = First interviewee from Mälarenergi
P4 = Second interviewee from Mälarenergi

**1. J: Då börjar vi med, vad är era roller på organisationen?**

**2. P3:** Om vi ska börja med mig då, [P3], så jobbar jag med styrsystem och kommunikation kan man väl säga.

**3. P4:** Och jag heter [P3] och jobbar som nätplanerare. Så investeringsplanering, förstärkningar, nya [inaudible], all exploatering, hur vi utvecklar nätet. Mer elektriskt än styrmässigt kanske.

[...]

**4. J: Hur länge har ni jobbat med det ni jobbar med nu?**

**5. P3:** Här på mälarenergi har jag jobbat med det jag jobbar med nu sedan 2004, men jag har väl egentligen jobbat med den här typen av, eller angränsande områden sedan 95 tänkte jag säga.

**6. P4:** Sedan 2015, så drygt 5 år.

**7. J: Och är ni insatta i deep learning eller någon annan maskininlärning?**

**8. P3:** Inte något djupare insatt, man har väl sniffat på ämnet någon gång tänkte jag säga och ja vi har väl haft en del projekt här som angränsar till det, men det är väl lite granna machine learning kanske, än deep learning. Och anomalidetektering och den biten, att man spelar in en baseline och tittar på avvikelser osv.

**9. P4:** Jag anser mig vara marginellt insatt i ämnet, inte min hemmabana.

**10. J: Nej, som sagt vi har ju utformat intervjun för att ni ska ju inte behöva kunna deep learning, vi frågar lite i början som där vi ser ifall ni har spontant som kan vara liksom intressant, men sedan kommer det frågor som liksom, ja, tanken är inte att ni ska ha koll på deep learning. Så då, nästa fråga om ni använder deep learning på företaget för smarta elnät, men då gör ni inte det lät det som?**

**11. P3:** Nej, vi har väl haft projekt som angränsar till det där, det var typ [inaudible], lite granna och, eller kan man säga.

**12. P4:** Kanske.

**13. P3:** Inte deep learning i sin fulla bemärkelse om man säger så.

**14. P3:** Mönsterigenkänning har vi väl i någon form jobbat med, men det är väl inte samma som deep learning antar jag.

**15. J: Det var inte liksom neuronnät.**

**16. P4:** Nej, nej, nej, det har vi inte gjort

**17. J: Vet ni om deep learning används inom smarta elnät någonstans idag?**

**18. P3:** Man tittar väl lite på det från olika elbolagshåll, har inte KTH med Lars Norström något samarbete inom smarta elnät? Eller smarta skydd, tillsammans med Ellevio och Vattenfall till exempel, och tittar lite granna på där med dubbla effektflöden och så vidare. Och sedan så finns det väl, vad är det mer då. Kommer du [P4] på någonting på rak arm?

**19. P4:** Nä.

**20. P3:** Det är sådana belastningsprofiler då, det är väl också vattenfall och Ellevio som tittar på Stockholm där, de har väl något projekt där. För att man ska kunna titta på, prognostisera ett halvår, eller tio år framöver och titta på hur det ser ut framöver.

**21. J: Ja men jag tror det hade några sådana projekt. Ja som sagt de här frågorna i början är ju liksom bara, liksom tänker lite korta frågor så här se om ni har något intressant att säga, det är liksom ingen fara om svaret är nej på dem, det är bara ifall det skulle... Ja, och sedan nästa var ehm, kan ni komma på några ställen där deep learning skulle vara användbart för smarta elnät.**

**22. P3:** Inom en rimlig tidshorisont så skulle man ju kunna tänka sig att man använder det för, vi håller ju på att byter ut alla elmätare i Sverige och då kommer vi ju få tillgång till mätdata på kundnivå på kvartsförbrukning. Och nu sitter man ju och analyserar och dimensionerar nätet utefter belastningsprofiler som är framtagna på 70-talet, man skulle ju kunna använda deep learning för att kunna analysera, ta fram andra typer av belastningsprofiler eller prognostisera belastningar i nätet på ett mycket bättre sätt än vi gör idag. I vårt fall skulle det ju röra sig om typ en hundra-tiotusen mätpunkter som man ska övervaka och dra slutsatser ur.

**23. A: Kan jag fråga vad belastningsprofil exakt är?**

**24. P3:** Ja men belastningsprofil det är för din lägenhet det är att på morgon slår du på spisen och kokar kaffe och sedan så är det ganska lugnt på dan för då är du på jobbet eller i skolan. Sedan kommer du hem och så slår du på spisen på kvällen och kollar på TV, då säger den vilken förbrukning du har över dygnet. Och den är ju olika beroende på om du har en villa med fjärrvärme eller om du har villa med elbilar eller om du bor i en lägenhet eller... Och olika typer av verksamheter har ju olika typer av belastningsprofiler och de har ju vi som dimensioneringsunderlag när vi bygger elnätet.

**25. A: Ok.**

**26. P4:** Mm, men det ser ju väldigt olika ut från dygn till dygn också en helg då har du en annan belastningsprofil också så att säga, och sedan är det ju väder och vind som spelar in också.

**27. A: Så det är alltså efterfrågan över tid?**

**28. P3:** Sedan är det ju så att elnätet byggs ju, det är så fort vi lägger ner en kabel så ska det här ligga i fyrtio år, så då är det ju fortfarande kan vi låta, om vi får bättre data så kanske vi kan köra den lite längre, alltså att vi vet att belastningen är lägre än vad prognosen säger. Då behöver vi inte byta i förtid för att byta upp det. Å andra sidan kan det ju vara så också att man får inte ta för tung vikt om vi tar fram nya belastningsprofiler, för belastningsmönstret ändrar sig ju över tid. Så vi måste ändå dimensionera med jättegoda marginaler. På 70-talet var det ju ingen som dimensionerade nätet för elbilar, men vi sitter ju med 70-talsnät nu som ska hantera elbilar. Där finns mycket data att leka med och man skulle kunna jobba med sådant.

**29. J: Jo, ja det var ju något som de ofta tog upp när vi läste teorin om det. Sedan kan ni liksom med den kunskap ni har se några utmaningar som skulle kunna dyka upp när man använder deep learning för smarta elnät.**

**30. P3:** Det är ju, allting drivs ju ekonomiskt, och blir det dyra installationer, alltså det gäller ju att kunna använda den utrustning man har i väldigt stor utsträckning. För det blir fort väldigt dyrt om man ska mass- vad ska man säga då, om man placerar ut utrustning på väldigt olika ställen så blir det ju fort en stor kostnadsbörda som då kunderna får betala. Så att det gäller ju att hålla...

**31. P4:** Nytta mot kostnad.

**32. A: Du menar om man skulle installera deep learning på system hos kunder?**

**33. P4:** Ja varenda mätpunkt varenda styrbar punkt, allting du ska ut i nätet för att kunna nyttja, få data eller styra saker har ju en kostnad. Och idag så har vi inte, vi har ju varken mätpunkter eller styrmöjligheter. Nätet är ju dumt i 99 % av fallet.

**34. P3:** Mmm, sen är det ju det om man kommer till just att kunna styra, då blir ju kommunikationen för den utrustning man kan styra med den måste ju vara väldigt tillförlitlig då också och det driver ju också kostnader.

**35. P4:** Sedan all sådan här automation [som] är till för nätet får ju inte ha kritisk driftpåverkan. Nätet måste ju funka ändå. Så att man får inte förlita sig för mycket på att stora delar av nätet hänger på någon form av automatisering, utan valda punkter kan man ha och hantera över tid, men inte hela nätet över lång tid.

**36. P3:** Hela nätet är ju uppbyggt på ett sätt att om en del, en skyddsfunktion fallerar ska det alltid finnas en skyddsfunktion som tar det högre upp så att säga.

**37. J: Ok. Japp, och sen... Det problemområde som vi fokuserar på med uppsatsen är att det är, alltså deep learning är en black box så man förstår inte hur system når sina slutsatser, eller vad som händer där inne, eller det är väldigt svårt att förstå i alla fall. Så att vi kommer sedan fråga om liksom specifika problem som uppstår med det, men kan ni se några problem spontant om man förlitar sig på sådan teknik?**

**38. P3:** Ja asså man måste ju se till att man säkrar upp den rent IT-säkerhetsmässigt, för där är ju en stor utmaning om man börjar använda publika nät för att samla in så mycket data, och man vill ju också då ändå på något sätt vilja förstå den här svarta lådan vill jag säga, kanske på en högre nivå då, men det är ju säkert mycket... det är ju mycket som man måste ha hänsyn till, man måste ha mycket, man måste ju ha någon form av regelverk kopplat till vad den här får göra, så att det inte blir för stora kostnader, för får den där för sig att koppla om i nätet hur som helst så kan det få jättestora kostnader för oss. Och så här false positives och true positives och false negatives, det är sådant man måste liksom ta hänsyn till permitteringen, hur den får agera.

**39. J: Ja, när du säger kostnader, menar du rent ekonomiska kostnader då eller typ att den slösar el...**

**40. P3:** Nej, det är inte el, den är nog försumbar, utan det är mer att den agerar på ett sätt som gör att vi kan få en kostnad av att vi tappar distribution utav el.

**41. J: Ok. Då går vi in på dem så här, det var 8 fördelar som man uppnår gör en mer förklarbart så jag tänkte fråga liksom hur de faktorerna påverkar smart grids enligt er. Så då den första har vi tillit, alltså att man kan lita på att systemet beter sig som man förväntar och vill att det ska göra. Framförallt är det, brukar det vara viktigt där att det inte introducerar nya fel. Så att tror ni att det är en viktig sak i smarta elnät?**

**42. P3:** Alltså en utmaning som jag tror att man får, alltså när man implementerar AI, det är en lång resa, på en sådan här stor maskin som ett elnät och om man tittar på de förutsättningar som man petar in i den här automatiken, det måste man ju jobba med ganska hårt. Det är inte allt för sälllan vi gör omsektionering av nätet, nätet ser inte alltid likadant ut, så det blir alltid nya förutsättningar som man måste ta hänsyn till, när det gäller sådan här automatik. Och får vi fel på en kabel då gör vi en omsektionering och isolerar felet, då blir det en helt annan förutsättning för den här svarta lådan att hantera elnätet. Och det skapar en väldigt stor dynamik i hur den här måste agera, så att just med regelverk, hur, man måste ha väldigt klart för sig hur den här, vad den ska göra och vad den ska arbeta efter, där tror jag är en stor utmaning och tilliten till det här, det måste man nog bygga erfarenhet på.

**43. P4:** Men jag här tillit, jag skulle inte se någon jättestor fördel eller jag ser mer risk än nytta med att låta ett AI styra elnätet i de kopplingar vi gör och den felsökningen vi gör. För vi ska ju hålla nätet så statiskt som möjligt, och felfallen är ju väldigt sällan förekommande och vid fel så har vi personer som är där och arbetar, och det vill jag inte, det är säkerhetsaspekter, så då måste vi ändå göra manuella åtgärder. Det man kan tänka sig, det är ju att tillåta vissa skyddsfunktioner för till exempel överströmmar och sånt där att vara dynamiska beroende på vilken omgivningstemperatur och lastprofil man har. Sedan om det skulle vara i ett AI, deep learning, för att använda det eller någon annan form av dynamisk last- vad heter det, dynamiska skyddsfunktioner, men det vet jag inte heller om det är AI eller om man vill definiera dem utefter ett regelverk. Så jag är otroligt skeptisk till att släppa loss någon AI till att styra själva nätet, analysera nätet, däremot det kan jag tänka mig.

**44. P3:** Ja, det är nog hursomhelst ganska långt fram innan man vågar sig på någon form av styrning skulle jag tro.

**45. P4:** Om du har en gubbe som byter skruvar så vill du inte ha en dator som du inte vet vad den gör, om den kanske [inaudible] strömmen.

**46. A: Kan jag tolka det som att ni tycker att tillit blir betydligt viktigare om AI-systemet är ute och styr elnätet medans om det till exempel gör prognoser så är det inte lika kritiskt?**

**47. P3:** Prognoser kan den få gör hur mycket som helst nästan, för det kan man göra rimlighetsuppskattningar från annat håll. Det är nog snarare bra hjälp att se mönster och sedan så kan man: oj, här var något intressant, då kan man verifiera det på samma sätt, så där sitter det mest en stor datamängd som behöver analyseras. Så där är det positivt, men själva styrning är jag mycket mer restriktiv till.

**48. A: Ja, med styrning menar ni då kopplingar och sånt, för det finns också ett annat sätt att styra på det är att styra värmeelement i hus så att elen används vid tillfälllen när det inte belastar elnätet för mycket. Är det något som skulle kunna vara.**

**49. P4:** Ja men det är flexibilitet, nu är ni inne på något helt annat. Det är ju hur du styr belastningsprofilerna och då är frågan vems brist är det du ska kompensera för, är det vår brist nätstationen? Är det bristen i mottagarstationen? Är det brister i regionnätet? Är det bristen i kapacitet på stamnät? Eller är det bristerna i produktion på grund av att det inte blåser? Så här finns det himla många olika nivåer, som styrs av olika personer, eller olika organisationer. Och att ha ett AI då som ska spänna över alla olika brister, är inte praktiskt genomförbart idag. Så jag kan lösa mitt problem i en slinga i ett villaområde med en sådan funktion, men kostnaden för det kontra mot att gräva lite grövre kabel är nog, ja, jag gräver nog hellre kabel, för blir det för kallt, då rycker man internetsladden och så kör man i alla fall. Och då kommer säkringen att hoppa, så att det här... Det finns ett antingen så gör du dig för beroende av AI:t, eller så finns det sätt att koppla förbi det, då håller inte elnätet, då är det lika farligt i alla fall. Så att vår lösning är snarare att dimensionera upp i sådana fall.
**50. A: Men om det är så att elnätet har problem med att leverera el för att det inte produceras tillräckligt, kan man, om det blir en peak av efterfrågan, kan det vara av nytta att reglera belastningsprofilen om det inte produceras tillräckligt mycket el för att gå runt liksom.**

**51. P4:** Nä men det är inte ett problem för oss. Asså att ett kärnkraftverk eller att producenterna inte har tillräckligt med el det är inte ett problem för oss på elnätsbolaget, på lokalnätsnivå. Så varför ska vi styra på det? Å andra sidan så kan det finnas hur mycket produktion som helst, men att det blir för mycket belastning i slingan hos oss för att vi har för många som har köpt elbilar i ett kvarter, så det är två helt olika problem som tar sig samma uttryck. Så vem ska vi styra på? Och det är därför jag är lite skeptisk till att lägga in för mycket intelligens i att styra värmepumpar och elbilar och annat. Jag tycker att man ska hålla, ja hålla ner man ska styra över laddning till exempel till natten vi nationellt regelverk istället. Man bör säkra mer hus och sedan lägga mer automation i enskilda hus att hålla nere effekten.

**52. P3:** Man kan ha ett litet lokalt energilager i huset eller i anslutning till ett bostadsområde kanske.

**53. P4:** Det är mer värt än att lägga det på en högre nivå med massa styrning och AI som kan styra åt fel riktning.

**54. J: Ja...**

**55. A: Jag vill säga bara, fick vi svar på tillit? Ni trodde inte att det var så relevant att använda AI för styrning och, men det kunde vara användbart för prognostisering, men där är inte, men kan jag fråga är tillit, hur viktig, hur relevant är tillit i den aspekt där AI är användbart, alltså hur viktigt är det att systemet fungerar som förväntat när det används på det sätt som ni tror är användbart?**

**56. P4:** Om vi bortser från styrning, då får man ta det efter vad man kan få, det beror ju på vad den bevisar. Jag menar har du ett bra AI som ger bra prognoser, ja men då ökar ju tilliten och även kravet på tillit jag menar om vi använder det för att det ger så bra siffror.

**57. A: Så det kommer av erfarenhet?**

**58. P4:** Ja precis.

**59. P3:** Men i styrning, om man ska se det som att styra, då måste det vara

**60. P4:** hundra procent.

**61. P3:** hundra procent [inaudible].

**62. A: Då kanske man behöver annat än erfarenhet, behöver man då validering på något annat sätt isåfall tänker ni?**

**63. P3:** Ja, mycket utav om man tittar på skyddsutrustning och sånt som när provar... När man provar en station till exempel som man distribuerar el med så blir det ju, man provar och provar och provar och provar tills man har provat igenom allting kanske två gånger. Så att man släpper inte en anläggning som inte är helt genomprovad över huvud taget, utan man måste vara helt tvärsäker på att allt fungerar som det ska.

**64. J: Ja, jag kom på när jag introducerade det här kanske jag råkade använda ordet "fördelar", det är lite missvisande, det jag menade var att det är fördelar med liksom förklarbar deep learning kontra vanlig deep learning, så att det är liksom sånt som kan bli problem med deep learning för att det är svårt att förstå det. Lite det vi var inne på att det blir problem med tillit när man använder deep learning. Men så det var väldigt bra svar tycker jag. Ja, och sedan så nästa faktor då är kausalitet, som innebär att om man förstår systemet så kan man liksom se de kausaliteter som systemet hittar. Tror ni att det är något som kan vara viktigt här. För att**

**liksom, eller ja så här, ett deep learning system, de hittar ju massa mönster så här orsak-verkan samband, då, det är ju svårt att liksom se vad de sambanden är om man inte förstår hur systemet funkar, så tror ni det skulle vara användbart att se de sambanden?**

**65. P3:** Ja asså det kan ju vara intressant att se vad den har byggt erfarenheten på så att säga, så man kan se att det finns relevans i det den har byggt upp beslutet på.

**66. J: Ja, ska vi gå vidare eller?**

**67. A: Ja, ehm, jag funderar på om, alltså ni tänker att om man förstår kausaliteten så kan man förstå hur beslutet har fattats, är det så du tänker?**

**68. P3**: Ja alltså så att den inte använder en helt ovidkommande del. Alltså det kanske uppstår någonting som kanske inte har med saken att göra. I samband med att en händelse sker.

**69. A: Så du tänker att man kan upptäcka eventuella fel genom att se orsakssambanden?**

**70. P3:** Ja, när man ser hur den har beslutat utifrån en viss händelse så kanske det går att [inaudible] mer av vissa delar som man använt som beslutsunderlag.

**71. A: Ja, jag vet inte om, vet du det Jesper, om [det som vi var ute efter här] är orsaken till beslutet eller om det är orsakssamband som den upptäcker, inte i sitt eget beslutsfattande, utan i världen?**

**72. J: Ja det vi mer är inne på är att man liksom kan hitta kopplingar ute i världen som systemet har gjort. Vi kommer in på det sen att man kan liksom mer verifiera och så.**

**73. P3:** Alltså man bygger sig en stor kunskapsbank om det man använder på flera ställen kan samköra data, där kan det väll finnas en fördel så att säga, eller vad?

**74. P4:** Jag sitter och funderar på vad jag tycker och tänker men om man då tar belastningsprofilen i vissa villaområden i stan så skulle man kunna dra slutsatser om vilken lönenivå folk ligger på i det området och hur tillväxten på elbilar kommer ske. Så det finns ju massa samband beroende på vilken data man har stoppat in.

**75. A: Det är det vi tänker på tror jag mer specifikt [...], så skulle det vara användbart att kunna se dem sambanden i konsumptionsmönster och sånt?**

**76. P4:** För den typen av slutsatser så är det ju intressant att veta varför vi borde satsa på att förstärka vissa områden före andra områden till exempel i prognoshänsende. Sen om vi har tillgång till den datan för att göra den typen av analyser det vet jag inte, jag tror man trillar på lite GDPR-frågor.

**77. A: Det är mycket möjligt.**

**78. J: Ja, men ska vi gå vidare då?**

**79. A: Jag tycker det.**

**80. J: Då har vi överförbarhet, det innebär att man kan använda kunskap som man har lärt sig från att lösa ett problem till att lösa liknande. Till exempel om man har ett system som kan känna igen personbilar så kan man använda kunskap från det för att känna igen lastbilar. Tror ni det skulle kunna vara något som är intressant här?**
**P3:** Alltså personbilar och lastbilar, tänker ni el…

**81. A: Nä, det är bara ett exempel som inte har med elnät att göra utan det är bara ett exempel i största allmänhet.**

**82. P3:** Ja alltså det var ju det [P4] var inne på tidigare att man kan ju dra lärdommar ifrån snarlika fall så att säga när man planerar något annat framtida område till exempel.

**83. P4:** Det finns ett område till man skulle kunna göra, om man tog alla skydd som vi har i alla stationer och sen lät dem kommunicera med ett AI så skulle ju den kunna lära sig hur utgångar beter sig och kanske även dra slutsatser och prediktera att fel kommer att komma. Frågan är om man kan överföra så mycket data med rimlig insats att AI kommer dra slutsatserna eller att man måste göra det på skyddsnivå. Men där skulle man ju kunna ha att en typ av linje skulle kunna avvika från en annan i sitt mönster men man skulle ändå se att om jag analyserar på det här sättet så kan jag göra det likvärdigt på de andra typerna av skydd så ja, om man försöker hitta på någonting så kan man se att det finns sådana fördelar med överförbarhet men det är, ja vi hittar på nu.

**84. J: Så det låter som att om det finns liksom någon vikt vid det så är det inte särskilt starkt, det kan vara bra men det är inget…**

**85. P3:** Det kan ju vara för att identifiera materialfel eller något sånt där att om man nu använder det här, till exempel ser att vissa värden sticker iväg och det beter sig på samma sätt så kan man identifiera att nu är det pågång här borta också, att något håller på att gå sönder. Alltså om [inaudible] till exempel och den beter sig på ett specifikt sätt så kanske man kan se och dra lärdomar utav... att det kommer att hända någonting här ganska snart.

**86. J: Nästa faktor, vi har varit inne lite på det här men det är informativitet, informativeness på engelska, som innebär att man får information kring hur beslut har fattats som man kan använda för att lära sig mer om det. Dels kan man granska beslutet och dels kan man lära sig av det. Till exempel i den artikeln vi har läst som tar upp det så har dem som exempel att en dokorstudent som frågar "var ska jag publicera min rapport" vill inte bara veta "här ska du publicera den" utan vill veta varför olika publikationer är bra så man kan använda det man själv vet också för att fatta ett informerat beslut. Vi har ju varit inne och nosat lite på det men är det något mer ni hade velat säga där?**

**87. P3:** Det var väll där egentligen det hamnade det där som jag, att det ska vara vädligt välbeskrivet, dels vad man vill uppnå med den här funktionen och sen är det ju så att man får ju inte bättre data ut vad man skickar in. Får man in dålig data så blir det ju inte någon bra data som man får ut heller. Så att det är ju sånt man får titta på när man bygger upp det där. Sen är det säkerhetsfunktionerna och vilka regelverk som man ska arbeta efter för att man ska nå en bra information utifrån det här.

**88. P4:** Om man tar ett annat exempel på det där, om man skulle låta analysera vilka ledningar som är hårt belastade så kanske den larmar för att den här är påväg att bli överbelastad, då skulle man ju vilja veta varför säger den det. Om det är trenden, om den alltid har varit det eller om det är en ökande trend eller, så att man kan, ja. Ju mer information om vad det är beslutet grundar sig på gör det ju lättare för oss. Då slipper vi ju ta reda på det själv.

**89. A: Kan man ställa frågan på ett annat sätt, omvända den lite. Om man skulle säga, skulle det gå att använda ett använda ett AI-system där man inte får reda på hur beslutet har fattats, alltså vad är det som ligger till grund för beslutet. Om man inte kan få reda på det, är systemet då fortfarande användbart för smarta elnät?**

**90. P4:** Nu är det en sån där öppen fråga igen, vilken applikation är det vi tittar på? För om jag tänker på analys av belastningsprofiler. Om jag säger att det här är överbelastat, ja då är det väll det. Kan jag

få mer information så är det bättre. Annars är vi på samma läge som idag, jag ser att det går mycket ström i kablen så får jag dra egna slutsatser.

**91. A: Skulle du kunna se nytta av det ändå av ett sådant system på något särskilt problem?**

**92. P3:** Rätt vald applikation eller rätt valt användingsområde men så fort man går in på skyddsfunktioner och stabilitet och sånt där då är det ett big no no. Då måste man veta att det är rätt beslut man fattar.

**93. J: Ska vi gå vidare? Nästa är rättvist och etiskt beslutsfattande. Det handlar framförallt om när man fattar beslut mot privatpersoner så vill man, dels kan det smyga sig in via algoritmen, antingen på träningsdatan eller hur den är designad som gör att vissa diskrimeras av den. Eller så kan den fatta oetiska beslut för att det är något man tjänar pengar på. Är det något ni tror kan uppstå här?**

**94. A: Ja kan ge ett exempel, om deep learning-system till exempel skulle upptäcka genom data att någon va alkoholist så skulle den kunna ge anonser på alkohol till den personer, för att ta ett exempel på vad det skulle kunna handla om.**

**95. P4:** Men om vi liksom pratar om leveransen av el så ser jag inte att det skulle kunna vara diskriminerande eller oetiskt. Kravet är hundra procent 24/7 och målet är att hålla sig över vattenytan hela tiden.

**96. P3:** Men om en sån här intelligens får för sig att, säg att man hade kommit mycket mycket längre, långt in i framtiden, när man börjar använda det här för att styra el att det är någon som inte förbrukar så mycket och den här väljer att ta in någon som förbrukar mer el, hur skulle man se det då, då blir det ju lite oetiskt.

**97. A: Kan du utveckla det jag förstod inte riktigt, om man…?**

**98. P3:** Om den här får för sig själv att… En kund är ju en kund för oss, det spelar ingen roll om den kunden förbrukar lite eller mycket. Och om den här ser då, om den har en parameter till att vi tjänar mera pengar på att ta in en som förbrukar mera el fortare.

**99. P4:** Nä, det kan den inte ta ett sånt beslut.

**100. P3:** Nä för den får inte ta ett sånt beslut, nä. Men om den får för sig att göra ett sånt beslut.

**101. P4:** Men då gör ju inte den sitt jobb, och då är det bara att slänga ut den på en gång.

**102. P3:** Ja ni va med på att det va en diskussion här bara?
**A: Ja jag förstår att det är ett hypotetisk framtida scenario.**

**103. J: Det som kan vara problemet är ju att med deep learning så vet man inte alltid varför den fattar sina beslut så då kanske man inte känner till att den fattar de här besluten som man ska slänga ut den för.**

**104. P4:** Det kommer såna här om du börjar använda den för investeringsplanering men det är ju inte där vi pratar just nu.

**105. P3:** Men om man använder en inparameter till att få tillbaka så mycket effekt som möjligt ut, om det är en inparameter så kan ju det bli ett problem.

**106. P4:** Mm, så mycket effekt så många kunder som möjligt i en störning men det är ju fortfarande vi har ju ett lagkrav på 24 timmar så... Vi ska i varje fall inte bryta mot lagen.

**107. P3:** Nä. Alltså, beroende på om man får ta till vara på för data där så tror jag man behöver granska beslutunderlaget på vad den baserar sina beslut på.

**108. J: Ok, ska vi gå vidare då?**

**109. P3:** Mm

**110. J: Jag vet inte om vi tänkte på att nämna det men vi har ju inte velat göra antaganden om vad som är relevant utan vi har ju velat fråga er om dem faktorer som gäller genererellt för deep learning. Så att ja, då går vi vidare till tillgänglighet som innebär att folk som inte är insatta i systemet har lättare att förstå det. Är det något ni tror är viktigt Eller att folk som inte är experter kan förstå.**

**111. P4:** Det beror på vad man använder det till men det som, för daglig drift så får det inte vara så. Jag vill inte att det ska… vi ska kunna leva utan det. Vi får inte vara beroende av det för då har vi satt oss i en sån sits att... För analys, ja, men då är det inte driftkritiskt heller.

**112. A: Alltså var det nödvändigt att andra än experter förstår systemet?**

**113. P4:** Nä.

**114. P3:** Inte för analys.

**115. P4:** Inte för analys och inte för… i och med att vi inte… jag säger att vi inte ska använda det för drift. Inte kritiskt drift. Och då kan man ringa in en expert när som helst. Du klarar sommaren utan systemet.

**116. A:** Man ska inte bli beroende av det.

**117. P4:** Nä

**118. J: Nä men för att andra har påpekat att man skulle kunna ha något halvautomatiskt system för drift, där lixom något operatör får rekommendationer av ett deep learning-system är det något du tror skulle kunna fungera?**

**119. P4:** Ja, det skulle man kunna ha. Alltså, aktuell belastning på vissa.. eller beslutsstöd men det ska fortfarande inte vara driftkritiskt.

**120. P3:** [inaudible] ...har tittat lite granna på det är ju också lite baserat på det

**121. J: Ja, då går vi vidare. Näst sista som är interaktivitet vilket innebär att slutkunder kan interagera lättare med systemet om det är lättare att förstå, är det något ni tror kan vara relevant?**

**122. P3:** Slutkunder? Det hamnar ju liksom lite utanför våran del skulle jag säga, eller hur ska man se det?

**123. P4:** Alltså i driften ska kunden inte ha någon påverkar på hur elnätet fungerar.

**124. P3:** Nä, men deras kostnader kan ju vara så ju vara så att de kan påverka resultatet men frågan är inte hur Google och sånt tar hand om den biten... att nu är det dyrt att förbruka mycket el så nu väntar jag med att värma upp mitt vatten till klockan 9 på kvällen istället då det är billigt. Och det tror jag blir andra aktörer som tar hand om den delen så att säga, vad tror du [P4]? Det här är ju min egen syn så att säga, det är ju inte Mälarenergis syn, utan det är min syn.

**125. P4:** Alltså för själva driften, nej. Och för kundbeteende, ja kanske men inte i huvudsak via oss då utan då kanske det blir en annan aktör på en högre nivå, en elhandlare.

**126. P3:** Vi ägnar oss ju åt drift så vi håller inte på med elhandel alls.

**127. P4:** Vi får inte hålla på med det.

**128. J: Ja, sen sista grejen va integritetsmedvetenhet. Som ja, här handlar det ju igen om privatpersoner så jag vet inte om det är relevant men alltså att systemet kan råka kränka folks intgeritet utan att man vet för att den drar slutsatser om den informationen den har.**

**129. P3:** Får inte ske.

**130. P4:** Det är ju vad den kan dra för slutsatser av våra kunddata då... peka ut elskurkar.

**131. P3:** Ja, där kanske kan finna men det finns ju andra metoder till att titta på, där det försvinner el så att säga ifrån det som vi mäter. Det som vi får betalt emot. Men det är också på elhandelsidan på ett vis men där är vi drabbade också.

**132. P4:** Svag koppling där.

**133. A:** Man kan ju tänka sig att med den datan man får in om man har 15 minutersvärden, man kan se ganska mycket vad en person gör kanske. När den går till jobbet och när den kommer hem och så vidare. Är det något ni tror kan vara risk att det missbrukas eller att det på något sätt kränker integriteten?

**134. P3:** Den informationskällan blir ju skyddsvärd så att säga. Det är ju ingenting som får passera ut så att tjuvar och sånt kan dra nytta av det.

**135. P4:** Alltså enskilda kunders belastning, den har vi ju koll på idag och den får ju inte spridas, för om vi vet att en viss kund har ett visst beteendemönster så kan ju den vara mer eller mindre attraktiv för en viss elhandlare och det får ju inte vi sprida. Men vi jobbar ju på vår nivå så det finns ingen integritetskränkning eller annat där. Vissa dem badar bastu varje kväll, vissa gör det. Men jag ser inte att man kan dra några sådana slutsatser som på något vis kränker någons integeritet beroende på hur mycket el dem drar.

**136. P3:** Nä jag kan inte heller se någon koppling där faktiskt.

**137. P4:** Sen styrningen av själva elnätet, den är ju väldigt opersonlig den har ju inte med kunder... Ja, i ändan på trådarna finns det kunder men själva nätet som sådant är ju inte kopplat till person.

**138. J: Ja, men då var vi nöjda.**

*the conversation continues on topics that are not relevant for the study*

## 7.6 Transcription of Interview 4

J = Jesper Lundberg
A = Alexander Lundborg
P5 = First interviewee from Energimyndigheten
P6 = Second interviewee from Energimyndigheten

**1. J: Vilka roller har ni på organisationen?**

**2. P5:** Ja, om vi börjar med mig då, jag sitter som senior rådgivare och fokuserar på området smarta nät och digitalisering. Och vad vi gör egentligen det är att jobbar mycket med forskningsfinansiering, utlysningar, publika medel så att säga för att främja forskning och innovation. Sedan så arbetar vi även mycket med omvärldsbevakning i olika nätverk eller med olika typer av uppdrag.

**3. P6:** Ja, och jag sitter som, vi kallar det väl för forskningshandläggare, så som Senja sade då när vi finansierar forskning och innovationsprojekt inom energiområdet. Och mina områden är ju då också smarta elnät och sen så har jag suttit en hel del med Senja med digitalisering och även med smarta städer.

**4. P5:** [P6] och jag jobbar väldigt tight ihop och det är jätteroligt

**5. J: Ok, hur länge har ni jobbat med det?**

**6. P5:** Om jag, [namn] säger då först så... Jag var ju med redan på tiden med smart grid Gotland med Vattenfall och sedan har jag jobbat som affärsutvecklare på ett start-up... Vad ska man säga med innovationsprojekt och start-up företag [unclear, probably InnoEnergy], så att jag vet inte hur lång tid tillbaka man ska säga men jag har jobbat inom energibranschen i ungefär 15 år. Innan dess var det på Eriksson med digitalutveckling av smartphones och liknande.

**7. P6:** Och jag, [namn] då, jag har varit på Energimyndigheten i ett år då, i den här rollen som forskningshandläggare. Och innan det så jobbade jag som konsult för Sweco och då med IT-system och IT-processer hos energibolag, så som elnät och flera värme...[inaudible].

**8. J: Hur insatta är ni i deep learning eller annan maskininlärning?**

**9. P5:** Oo, den är bra! Ja, vi har ju koll på vad vi kan göra med olika teknologier och med data och asså med artificiell intelligens och så vidare, men vi är ju inte insatta i detalj i hur saker och ting fungerar, utan det är så att vi kommunicerar med KTH och alla möjliga olika forskare, folk som jobbar inom institutioner och liknande inom de här områdena, men vi är ju inte experterna på just deep learning så att säga. Då skulle vi inte sitta på de här positionerna, däremot har vi kontakt med folk som jobbar mycket med den här typen av frågeställningar.

**10. P6:** Och vi har ju haft en utlysning då nu där vi har eftersökt projekt som undersöker hur artificiell intelligens, hur man kan använda det för att, vad ska man säga, för att gå mot ett mer klimatneutralt energisystem.

**11. P5:** Ni får ju komma ihåg att vi är ju liksom myndighet, så vi kommer ju inte, vi styrs ju av vad vår uppdragsgivare, det vill säga regeringen, vårt departement vill, så att säga. Och utifrån att de ger våra energipolitiska mål som styr vad vi gör för någonting. Och just nu är det omställning som vi är väldigt aktiva inom, energiomställning naturligtvis. Det är där alla jobbar inom energibranschen. Och sedan så naturligtvis klimatmålen.

**12. J: Ja och som jag nämnde när vi mailade så tanken är ju inte att ni ska liksom behöva ha någon särskild koll på deep learning utan intervjun är ju utformad så att det liksom, ja, det ska funka oavsett.**

**13. P5:** Det låter bra, jag tror vi kan svara på det utifrån de frågorna, men som sagt var, experter är vi inte.

**14. J: Bra. Vet ni om deep learning används inom smarta elnät någonstans idag?**

**15. P5:** Ja, det beror ju på exakt vad definitionen på deep learning är. Om vi börjar med artificiell intelligens. Fortum tittar ju en hel del på det på fjärrvärmesidan. Sedan så är det vi pysslar ju mest med forskning och innovation, så vi har forsknings- och innovationsprojekt som olika demonstrationsprojekt, olika prototyper, olika typer av analyser av att använda de här teknologier, så att säga, så att det är ju data och hur man använder data på ett smart sätt är ju högaktuellt. Och det är ju precis som [P6] sade, vi har precis haft en utlysning inom området också.

**16. P6:** Jag vet att man har tittat på det i vissa elnätsbolag då att använda det för att kunna prognostisera, alltså kunna göra bättre prognoser av elanvändningen i nätet för att kunna. Nu är jag lite osäker på eran bakgrund, kan ni, har ni elnätsbakgrund också?

**17. J: Nej, alltså vi har ju mer bakgrund i IT, men vi har läst på lite om elnät för den här uppsatsen, men det är inte vad vi har vår bakgrund i liksom. Så vi är ganska amatörer där.**

**18. P6:** Kapacitetsbrist, säger det er någonting?

**19. A: Alltså att det inte tillverkas tillräckligt mycket el för att ja, det produceras inte tillräckligt med el, det finns en brist på el som produceras antar jag?**

**20. P6:** Mer, vad ska man säga, när man pratar om det inom elnät så är det mer att möjligheten att överföra, så att oftast finns det tillgång till energi men sen så om man tänker sig elnätet som motorvägar så kan det ju bli kö ibland in i storstadsområden för att det är många som vill använda det samtidigt, så det är väl kanske skulle man kunna säga. Kapacitetsbrist alltså att man inte har, att det är många som vill använda el samtidigt på samma plats, och då har man inte utrymme för det i elnätet, och det är en högaktuell fråga för elnätsbolagen just nu då. Och då vet jag att det finns elnätsbolag som för att kunna prognostisera och också liksom veta hur man ska bygga ut näten framåt, så vet jag att man har tittat på machine learning för att kunna göra bättre prognoser, sedan är jag osäker på om, jag tror inte att det är deep learning och jag vet inte om det är machine learning heller, men jag vet att man ofta hos elnätsbolag tittar på att använda data för att kunna göra predictive maintainance, alltså att kunna förutse, att man ska kunna förutse underhåll, för tidigare så har man mer jobbat med att man underhåller stationerna, och byter ut komponenter i stationerna utefter en rådgivning från den som har levererat de här komponenterna. Så säger man så här, den här håller ungefär... Man måste byta den här komponenten inom 10 år och den ska underhållas inom 5 år. Men det man kan göra istället är ju liksom att mäta, det beror ju på komponent då, men man tar upp massa mätdata från de här komponenterna och sedan så kan man liksom via mönster se till exempel om någon komponent havererar så kan man liksom se skiftningar i de här mätningarna och något värde går upp eller går ner så här, och så istället kunna förutse att.. eller med den här datan då kunna förutse istället när man istället behöver byta en brytare istället för att byta det var sjunde år så kanske man kan ta den här datan och förutsäga att just den här komponenten måste vi byta nu, men de här andra komponenterna de kunde sitta i tio år till. Så att för att förutsäga underhåll vet jag också att det används.

**21. P5:** Mmm, det har också att göra med hur smarta till exempel basstationer är och hur mycket mätvärden och sådant som kommer ur saker och ting så att säga, så att precis som [P6] säger, det går åt det hållet, och man har börjat med det, jättebra exempel [P6].

**22. P6:** Sen har vi ju också, om jag bara, eller jag kan nämna att vi, jag vet också att vi har ett projekt som vi har finansierat där man ska titta på eller man tillämpar deep learning-metoder på stora datamängder i el-kraftssystemet, och då så mäter man spänningar och ström och så försöker man titta på, eller i närheten av förnyelsebara energikällor så som sol- och vindkraft och så försöker man utifrån de här datana [sic, presumably meant data as determined form plural] kunna hitta orsaker till störningar, för om man kan se att de här förnyelsebara... får se nu så jag säger rätt. Ja, men i alla fall att kunna göra de här mätningarna på spänningar och ström och för att kunna identifiera mönster för störningar i närheten av vindkraft och solkraftsparker.

**23. P5:** Jag kan tillägga också, det stämmer [P6], att underhåll också när det gäller vindkraft som kanske är ute till havs då, som kanske inte är så otillgängliga [sic, presumably meant "tillgängliga"], det är också ett område man tittar på. Så att på produktionssidan tittar man ganska mycket på underhåll och hur man ska kunna ta bort driftstörningar egentligen och hur man ska kunna prognostisera vad gäller väder och vind och liknande och för att effektivt utnyttja produktionsanläggningarna så att säga. Sedan så tittar man på det [P6] var inne på, kapacitet och laststyrning, och hur liksom flexibilitet, hur ska liksom. När vi inför en massa, tidigare hade vi baslast, mycket kärnkraft och mycket vatten, det har vi fortfarande, men nu har vi liksom mer och mer gått in på att vi även har förnyelsebart som är väldigt varierande beroende på väder och så att säga beroende på hur mycket man producerar kan ha att göra med hur mycket solen lyser till exempel då, sen har vi lokala aktörer, så det blir många spelare i något som tidigare var ganska vad ska man säga, en lina som matar ut allting som var ganska beräkningsbart. Nu är det mycket som händer i den här miljö... omvärlden så att säga. Många olika producenter och... både lokalt, kan va i någon energigemenskap, det kan vara någon som har solceller på taket samtidigt som det kan vara stora vindkraftparker som producerar när vinden blåser, det kan vara. Ja i samband med att vi avvecklar kärnkraften så blir det mer och mer liksom ett nät som måste styras på ett annat sätt, och då kommer mycket av de här teknologierna in. Och då har vi även marknadssidan och handeln med det hela, och handel och flexibilitet är något som kommer ganska mycket så att säga. Ja, jag har så här mycket över och jag kan tillhandahålla besparingar på el om ni behöver elen någon annanstans och den typen av affärsmodeller så att säga. Jag vet inte om jag lyckas förklara det logiskt, men det finns ju de som kallas aggregatorer, som i nätet så att säga som kommer mer och mer som tillhandahåller till exempel en fastighetsvärd, ta Örebro bostäder som har kommit ganska långt i det här, de kan till exempel minska värmen en aning i sina fastigheter så kan man på så sätt tillhandahålla mera energi till exempel. Eller man kan stänga ner en stor fläkt eller någonting sådant, kanske ett litet tag bara för att hjälpa elnätet så att säga. Att klara av det hela när man har väldigt stor förbrukning just då. Det kan också ha med priser att göra, till exempel prissättning och prissignaler. Så det är väldigt viktigt område, hur man både hanterar data, tar smarthet, intelligens i det hela och att kunna prediktera olika saker och ting. Ja, vi hade ju projekt, Gotland är ett exempel, Coordinet, till exempel. Är ju ett EU-projekt just nu som pågår där vattenfall är inblandat. Fortum har mycket inom fjärrvärme och i södra Sverige är det marknadsplatser också, kraftringen är inblandade i det. Så att det pågår en hel del. Kring Uppsala är det väldigt spännande också, som [P6] var inne på med kapacitetsbristen, att man har ont om energi vid vissa tillfällen men man har höga toppar. Och då är det också spännande och det går mer och mer också att ha små energigemenskaper eller mikronät så att säga, hur de kan bidra med, den typen att man stänger av någonting, det kanske blir ganska lite i varje hushåll eller varje lägenhet, men det kan bli ganska mycket om man ser till att samla ihop det hela så att säga. Och har många avtal med olika typer av aktörer.

**24. J: Det var ju intressant.**

**25. P5:** Förlåt, jag kan bara lägga till något, om man tänker på nationell nivå så har vi ju till exempel transmissionsnätet, där finns det ju olika typer av nät, men till exempel där är det intressant även längre ner på regionnäten till exempel, men avbrottshantering kostar väldigt mycket pengar till exempel, eller att säkerställa att ställa trygg energiförsörjning. Den typen av frågor är också intressanta att kolla på. Elkvalitet, att man måste upprätthålla frekvens i nätet för att det ska fungera så att säga. Och det är också viktigt. Där att kunna titta på de parametrarna, att få in det här och kanske kunna förebygga avbrott att kanske kunna simulera saker och ting till exempel, det kan vara cybersäkerhet.

Att kunna simulera till exempel angrepp för att förebygga dem, den typen också. Det finns mängder applikationsområden på det ni skriver om och det gör det så otroligt intressant.

**26. J: Mm, jo men det är det. Det kan ju också vara svårt att säga generellt när det är så mycket det handlar om men kan ni generellt se några utmaningar som skulle dyka upp om man använder deep learning inom de här områdena?**

**27. P5:** Jo det är väl... En sak är ju aggregera datan, men det är ju klassiskt, så det är inte något konstigt just för det här området. Det finns hur många utmaningar som helst, jag vet inte var jag ska börja, [P6] [inaudible]
**P6:** Men det som ni tittade på då var det som jag förstod, var det kopplat till att det var en black box, är det det som är…

**28. J: Ja alltså, vi kommer gå in på mer specifika grejer sedan.**

**29. P6:** Ok, så generellt, vad sade du nu, vilka utmaningar det finns?

**30. J: Ja**

**31. P6:** Juste, precis, exakt. Nu spekulerar jag lite, men en sak, men jag tror att... Ja dels är det väl kanske också att hantera mängden data, men det kanske liksom inte är unikt för elnätsområdet liksom, men jag tror att, det som jag funderade lite på precis som Senja var inne på att när man för till exempel så samlar man ju in massa mätvärden från elmätare i elnätet och kan göra många typer av mönster- eller liksom analyser på det, men just elnätsdatan är ju, är en känslig uppgift så då om man komma in på individnivå och komma på vart och vem mätaren sitter på och koppla det till vem som bor där så är det en känslig uppgift. Men kan man aggregera så att man säger att man inte kan identifiera vem det är, om man samlar data från många olika hushåll eller företag eller sådär. Då är det inte en känslig uppgift om man kan göra analyser kring det. Men det har ju visat sig tror jag liksom i andra områden har jag förstått att man ändå kan, även om har aggregerad data så kan man urskilja ibland individer, så det tänker jag kan vara... Till exempel så läste jag ett exempel om att man kunde identifiera vem... Typ om man skriver på ett tangentbord så kan liksom, kan man identifiera vem det är som skriver, för varje individ har ett mönster för hur man skriver och kanske hur snabbt man når till de olika tangenterna och sådär. Och kan man göra det med elnätsdata så tror jag att det kan vara... alltså om du kan börja urskilja individer ur den aggregerade datan, så att vad kan man säga, privacy, vad heter det.

**32. P5:** Integritet kan man säga tror jag, och den är viktig, den är jätteviktig.

**33. J: Ja, vi kommer in på det senare, det är ju intressant att höra att liksom ni tar upp det liksom opåkallat, det tyder ju på att det är viktigt, det är lite därför vi vill fråga generellt först.**

**34. P5:** Och sedan en annan fråga, nu hann jag tänka lite igen [P6]. Det var bra att du tog upp det. Interoperabilitet mellan olika typer av tekniker och mellan olika typer av lösningar så att säga, för allt ska ju, mycket handlar om att lösningarna finns, nu kanske jag tar i lite grann, men en hel del teknik finns där ute. Men det handlar om att koppla ihop dem och få dem att fungera tillsammans och dra nytta tillsammans i olika typer av systemlösningar. Och där har vi interoperabiliteten emellan olika typer av lösningar och få saker och ting att lira på olika nivåer tillsammans för att på något vis bli vettigt så att säga. Och där har vi också standarder, som också nu... Är det IT-standard, är det elstandard, hur hänger saker och ting ihop. Sedan så har du även lite det här att utformning av marknadsplatser, både regleringsmässigt och så att säga, hur ska det fungera med olika typer av lokal, ska det vara flera marknadsplatser, ska det vara en, hur ska det fungera med aggregatorer, hur ska det fungera med flexibilitet, med lager, den typen av tankar finns. Och man jobbar väldigt mycket med detta, både vi, energimarknadsinspektionen och en mängd andra aktörer. Sedan handlar det lite granna om vem ska ta investeringarna och risken, och när ska man investera? Säkerhet kommer automatiskt med brev på posten så fort vi pratar data. Och där har ju vi ett ansvar också utifrån MSB som

samordnar olika typer av sektoransvar tror jag det kallas. Men i alla fall där har vi som tillsynsmyndighet utifrån internationella regler om olika typer av sabotage, om avbrott, olika typer av... Att vi är en tillsynsmyndighet. Ja, ni förstår hur jag tänker i alla fall, utifrån ett säkerhetsperspektiv. Och utifrån att ha en trygg leverans, utifrån det perspektivet tittar vi på det hela så att säga. Sedan så finns det även lite tankar kring när man ska investera, vem som ska göra vad. Jag menar om vi tar smarta mätare, är det våran elleverantör som betalar mätaren hemma, eller är det vi som konsument som ska göra det, och vem gör vad egentligen? Och vem skrivs avtalet med? Och hur ser affärsmodellerna ut? Har jag, finns det något mer [P6]? Det är ganska stora puckar det här.

**35. P6:** Ja, nej men jag tror att det är, nej men jag tror absolut att du har fångat det.

**36. P5:** För att göra det lite enklare så har vi dels de som producerar elen, sedan har vi distributörerna, nätet, sedan så den rollen ändrar sig lite grann nu också för nu har vi lokala producenter som egentligen är konsumenter som till exempel matar in sin solel eller använder den för att försörja sig själv så att säga. Sedan kan du för att göra det ännu mer komplicerat ha små communities så att säga med egna små nät till exempel micronät av olika slag som också kan bidra till ett stort system. Och då blir det inte att elen matas från en central plats direkt ut mot en kund utan på en gång blir det att den går upp och ner samtidigt både lokalt, regionalt och även på nationell nivå. Och sedan näten är ihopkopplade också för att vi hjälper ju varandra med våra grannländer också. Och sedan så pågår en mängd handel däremellan då.

**37. P6:** Jag tror precis att om tittar på de stora hindrena för att tillämpa deep learning, eller "stora hinder", men ja men precis som Senja var inne på att som elnätsbolag så är det ju man har ett leveransansvar så det är väldigt... Det får inte vara några stora avbrott och det får inte vara några långa avbrott så det måste finnas en hög tillförlitlighet då för att tekniken fungerar. Och att det gör kanske, om man ska generalisera branschen lite, så gör det ju att man lite generellt är lite mer restriktiv till vilka nya tekniker man plockar in. För det kan inte bli, i princip så, eller ja, det kan inte bli svart, och framförallt inte till vissa samhällsfunktioner. Och blir det det så måste det vara under ett väldigt kort tag då. Man måste ha höga krav på sig att leverera.

**38. P5:** Samtidigt då som man ser möjligheter att spara kostnader med det här.

**39. P6:** Ja men absolut! Precis. Exakt.

**40. P5:** Sen så det blir mycket krav på att vara med i utvecklingen samtidigt som man har krav på att man kan inte göra hur som helst i ett elnät. Det är som du säger [P6]. Det egentligen leder till en sak till, det är kompetens, det är inte så lätt, en del ägs av kommunägda bolag till exempel, det lokala bolag som... Det är mycket förändringar, mycket som pågår, mycket... Ett helt paradigmskifte egentligen.

**41. J: Ja, det är väldigt intressant, det är också mycket som kommer in på det som vi har uppfattat som problem, så det är väldigt intressant att ni liksom, det verkar relevant. Sedan, vi kom ju in lite på det här med black box, men kan ni komma på några utmaningar som handlar specifikt om det?**

**42. P6:** Får jag, för black box om jag har förstått det rätt, det handlar om att man inte kan följa hur algoritmen kom fram till sitt svar?

**43. J: Ja precis.**

**44. P6:** För det är liksom indata-svart låda-utdata och så kan man liksom inte se vad som händer i den här lådan

**45. J: Ja. Sedan är det inte riktigt så liksom svart och vitt, men ja.**

**46. P6:** Ok. Ja.

**47. P5:** Ett problem är ju tillförlitlighet, finns det tro till systemet så att säga. Sedan så, ja det är väl en av de sakerna med black box, om vi talar om förståelsen, om vad som händer egentligen i det här nätverket så att säga. Sedan så tänker jag också på vad händer med alla signaler till exempel om du styr ett kärnkraftverk, ett vindkraftverk eller vad du styr för någonting så kommer det in en mängd signaler till en driftcentral som samordnar och styr alltihopa. Och jag menar någonstans när det blir fel måste du kunna ha spårbarhet. Eller i alla fall någon slags både tillförlitlighet och spårbarhet så att man vet vad som händer och kan styra på detta så att säga. Alltså om det gäller till exempel produktionsanläggningar. Gäller det andra så är det lite annat, men det måste fortfarande vara så att energimarknadsinspektionen [inaudible] roller och så att säga att det finns både säkerhet och att det finns roller när det gäller till exempel jag tänker på blockchain att det är tillförlitligt att man agerar som en krediterad aktör eller något sådant. Att det finns ett system som gör att man får den typen av tilltro till att det här är seriöst, det är lagligt, det följer marknadens regler, den typen av kan jag tänka mig.

**48. P6:** Men är det, för det är specifikt kopplat till elnätsbolagens verksamhet, den här frågan?

**49. J: Ja precis.**

**50. P6:** Precis och jag tänker också att det kanske beror på lite precis vart man tillämpar, nu spekulerar jag också, men vart man vill tillämpa artificiell intelligens. Till exempel, om man tillämpar det så liksom i hur man driftar nätet, och även för komponenterna i nätet, alltså det här underhållsperspektivet, när det behöver underhållas och sådär. Och om man kan förutse störningar och så, då kanske man till exempel vill ha högre tillförlitlighet, det vill säga att man vill kunna förstå hur algoritmen kommer, kom fram till sitt svar. Men sedan så skulle jag kunna tänka mig att det finns till exempel som om man prognostiserar hur det bäst är att bygga ut nätet på tio års sikt, det är klart att det påverkar investeringar och så men då är det kanske inte riktigt lika viktigt att förstå hur algoritmen kom till sitt beslut utan att man tar det som ett av sina underlag till beslut. Att "algoritmen kom fram till det här, det verkar rimligt, vi kör på det.", eller ja. Så att det beror nog också på inom vilket område.

**51. P5:** Vi spinner på här ganska bra [P6], jag tänkte bara dra en sak till, alltså det handlar ju också om att identifiera use casen för det är på olika nivå, jag menar en del saker kommer man vilja, jag menar risken är att Sverige blir svart eller delar av Sverige blir svart då kanske en nödlösning finns på plats för det ska bara fungera så att säga, så att för att få någonting att hända. Men i andra sammanhang kanske det är mer att man vill att systemet ska fungera mer som ett beslutssystem där man får underlag för att sedan kunna fatta ett beslut och sedan är det någon faktisk fysisk person som fattar det beslutet. För att göra si eller så, så att det handlar väl mycket om att i den här boxen att ha use casen så att säga i olika sammanhang, sen så även att, att ha tillgång... Vi har ju data i systemet, men på olika nivåer och kanske, vi har inte all data där som vi skulle behöva idag. Utan det handlar väl mycket om att bygga ut så man får tillräckligt mycket sensorer och tillräckligt mycket aggregerad data så man faktiskt kan fatta beslut beroende på vad man gör för någonting så att säga. Jag kan ju bara säga att till exempel jag bor i [område i Stockholm] i den delen som byggdes först. I min lägenhet finns det bara en enda sensor för att känna av värme, den är i hallen. Så skiner solen in genom fönstren så fattar inte hallen att det faktiskt är varmt här inne, utan den kör på elementet och det är väldigt varmt och gosigt här inne. Och då öppnar jag fönstren, och hur effektivt är det här? Så att det är bara för att ta ett väldigt simpelt exempel så att säga. Så att det handlar mycket om att ha sensorerna, bra data, att aggregera den. Sedan handlar det också om överföring naturligtvis, vad är bästa mediet att överföra, alla basstationer, till exempel var man nu har elnätet någonstans, det finns liksom inte så att det automatiskt är wi-fi eller att det är något slags internet utdraget dit så att säga, utan då kanske det måste finnas någon lokal smarthet som kanske drar lite slutsatser, man kanske inte kan, man kanske för över det via 5G-nät eller den typen av, så det måste vara integrerade typer av lösningar för

dataöverföring, någon slags bra aggregering och vi behöver tillgång till data. Och sedan så behöver vi integrera det i nuvarande system, det vill säga antagligen beroende på vad vi pratar om, men det måste ju fungera i ett fastighetssystem eller i ett Siemens, säg styrsystem av produktionshandläggningar eller vad det nu är för någonting. Och sedan så har vi det här med säkerheten också att det måste vara säkert. Och där har vi både hårdvarusäkerhet och mjukvara. Jag menar hårdvara är ju komponenter till exempel av olika slag som tillverkas i Kina, vad vet jag, eller liksom hur de fungerar och vad för backdoor, eller vad det finns för någonting där till exempel. Så att det handlar mycket om också vad vi sätter in i systemen och att vi inte bygger in oss i ett hörn så att säga.

**52. J: I litteraturen har vi upptäckt åtta teoretiska fördelar med begripliga AI-system, som alltså borde saknas i vanliga deep learning-system som inte är särskilt begripliga. Så då tänkte vi att vi går igenom dem, en efter en, och så förklarar ni hur relevant ni tror det är för smarta elnät, eller hur viktigt det är. Och många av de här har ni redan tagit upp så att jag tänker att jag går igenom ändå och så får ni säga om ni känner att ni har något mer att säga om det eller om vi ska gå vidare. Så den första är ju tillit, som ni har varit inne på rätt mycket redan. Är det något mer ni känner att ni vill säga där?**

**53. A: Vi kanske ska definiera vad vi menar med tillit.**

**54. J: Ja, juste. Det är att systemet beter sig som förväntat och det är framför allt viktigt att det inte introducerar nya fel som inte fanns tidigare.**

**55. P6:** Juste, ska vi säga hur relevant det är för?

**56. J: Ja eller hur viktigt är det?**

**57. P6:** Hur viktigt, ah. Ja men det skulle jag nog säga är viktigt för smarta elnät. Mycket viktigt.

**58. P5:** Jätteviktigt, jag håller fullständigt med. Vill du att vi ska gradera på en skala i slutändan eller vill du bara att vi ska säga viktigt eller vad vi tycker är mindre viktigt eller hur vill du att vi ska...

**59. J: Du kan säga, liksom, om ni tycker det är, hur viktigt bara i ord och förklara vad ni menar men det här har ni också pratat om rätt mycket så...**

**60. A: Får jag följa upp, kan man bygga upp tillit utan att förstå hur det fungerar? Alltså, om det är en black-box kan man då ändå få tillit för systemet tror ni?**

**61. P6:** Eh, teoretiskt sätt så ja, det tror jag.

**62. J: Handlar det om att man testar isåfall?**

**63. P6:** Ja det tänker jag, alltså att man testar. Precis, och det handlar väll då att få möjlighet att testa och bekanta sig med systemet och testa, alltså typ jämföra med verkliga fall eller vad man ska säga, alltså hur hade vi agerat utan det här och hur agerar vi med det eller vad man ska säga. Så då tror jag, jag tror att man skulle kunna bygga en tillit till systemet.

**64. A: Varierar det beroende på i vilken användningsområde man har det? Är det vissa användningsområden där det verkligen är nödvändigt att man förstår hur det fungerar och andra där det räcker att ha testat det på något sätt, men inte riktigt förstå hur det fungerar invändigt.**

**65. P6:** Ja men absolut, jag tänker att det är... precis så, olika användningsområden. Och jag tänker också att är liksom ändå också en teknik som är... precis, att i vissa områden så är det viktigt att veta och i vissa områden så kanske det är mindre viktigt vilket gör att man kan känna tillit utan att förstå hur det funkar, lite beroende på vad man ska använda det till. Sen tänker jag, nej förlåt jag tappade det. Säg Sanja.

**66. P5:** Vi tänker kanske på samma sak för det beror på lite grann vilken detaljeringsnivå man behöver ta... alltså, dels har det med risk att göra. Desto mer risk det är desto mer vill man veta att det är tillit så att säga, att... ska jag styra ett kärnkraft då fasiken ser jag till att jag liksom har koll på att det här liksom är testat och att det fungerar men man kanske inte behöver veta alla detaljerna alla gånger. Det räcker kanske med en konceptuell modell men det är också så att när någonting inte fungerar så måste man veta mera än bara det övergripande om hur det borde vara så att säga. Så mycket handlar om tillförlitlighet, att man ser att det fungerar, sen handlar det om risknivå lite grann, vad är det man styr för någonting och vad händer om man gör fel eller det går fel. En del saker som kanske ingen risk alls men ska bara automatiseras och dem vill man inte ens se men någonstans handlar det om att vad man än styr för någonting... jag blir väldigt irriterad om jag inte har en konceptuell modell över vad som, vad det är för någonting som systemet styr så att säga. Jag behöver inte alls veta detaljerna men jag vill veta att det finns ett värmesystem i huset och att det finns en central någonstans som styr det hela och att det optimeras på x antal, och jag kanske kan påverka det här. Alltså någonstans måste man ha en konceptuell modell det är väll det som jag tycker är ganska viktigt när det gäller tillit. Och sen så vilken nivå av risk.

**67. A: En konceptuell modell över hur systemet fungerar?**

**68. P5:** Ja, det kan ju va alltså... en bil, fyra hjul, motor... Jag menar, det sitter ett chassi du har ett bagageutrymme och [inaudible] eller så kan det vara betydligt mer detaljerat men det är lite beroende på hur mycket man själv, så att säga, känner man tillförlitlighet och vilken beslut man ska fatta också. Är det standardbeslut som [P6] var inte på lite grann då kanske man inte vill veta någonting om systemet man vill bara veta att det fungerar. Och det kanske inte spelar någon roll om det blir något fel någonstans för då vet man att då avbryter vi och så kallar vi på någon tekniker, och då är saken löst. Jag kan gå och ta en kopp kaffe extra. Men spelar det roll däremot så blir man ganska förbannad när man inte kan liksom göra någonting i systemet och det är väldigt bråttom och alla chefer ringer, då blir du ju jättestressad.

**69. P6:** Ja, precis så det är väll kopplat som [P5] säger till vilken risk, till exempel inom drift, alltså för att drifta systemet och också även inom vissa delar av underhållning då om man riskar att komponenter går sönder så att man inte kan leverera el. Inom dem områdena då blir det liksom en högre risk men om man ska göra analyser kanske på elanvändning av andra syften då kanske det inte är lika...

**70. A: Ja, är det prognoser och sånt eller?**

**71. P6:** Ja precis exakt det kanske mer är prognoser eller mer är... precis. För att skapa någon typ av förståelse för elanvändaren eller för att kunna implementera typ alltså såhär förnyelsebar produktion eller sådana där saker då kanske det är...

**72. P5:** Ja, då handlar det kanske mer om konsumenten. Jag som konsument kanske inte vill veta så jättemycket mer än att det styrs och det finns ett system och det finns en ansvarig och det finns ett telefonnummer jag kan ringa om det inte blir perfekt. Så kalibrerar dem systemet eller någonting sånt. Det beror ju på vad det är för någonting. Alltså jag tänker på värmen tänker jag på i det här fallet, det är skillnad om jag inte får någon el alls för då blir jag tokig. Eller optimering av olika tariff, prismodeller till exempel. Då måste jag på något sätt iallafall ha en aning om

någonting och sedan så kan jag ringa någon som... min, den som jag har avtal med eller någonting sånt som kan förklara lite mer i detalj, då är det helt okej. Då behöver jag inte veta så jättemycket för jag vet att någon hanterar det åt mig.

**73. J: Då går vi vidare tänker jag. Nästa grej är kausalitet vilket innebär att deep learning-system hittar en massa korrelationer och om man kan förstå hur systemet funkar kan man se vilket håll kausaliteten går i dem korrelationerna, alltså orsak-verkan-sambandet. Är det något ni tror skulle vara viktigt i smarta elnät, att kunna se vilket håll kausaliteten går.**

**74. P5:** En fråga, när ni säger så, hur ser ni på smarta elnät? För att, ser ni själva nätet, ser ni på produktionsanläggningarna eller ser ni det som basstationerna eller ser ni det på tjänsterna framför mätaren till exempel eller alltså ute hos konsumenter av olika slag, eller?

**75. J: Alltså vi tänker ju framförallt på elnätet men sen också det här att man kan använda smarta elmätare för demand response, alltså påverka elkonsumtion men det är framförallt själva nätet vi tänker på.**

**76. P5:** Och vilken roll pratar vi om här, pratar vi om någon som är distributör och styr nätet eller pratar vi om någon som är konsument som använder energi eller elen då så att säga eller hur ser vi liksom...?

**77. J: Det skulle väll framförallt vara den som styr nätet som är intressant. Och sedan så är det ju alltså, det är ju viktigt också för kunderna att det funkar bra liksom, för dem som använder elen men fokuset ligger ju på den som styr...**

**78. P6:** Jag tror att det är viktigt att förstå orsak... i styrning av nät, alltså som om man tänker dem som är nätbolag och styrning av nätet att man tycker det är viktigt att förstå och kunna följa sambanden. Nu har jag inga direkta exempel men det fanns, det finns ju något exempel inom sjukvården där man skulle ta hjälp av artificiell intelligens för att diagnostisera hjärntumörer tror jag, på röntgenbilder och när man sen såg hur algoritmen diagnosticerade så såg man att den hade identifierad ett datum som tydligen var i underlaget ganska vanligt att, vad heter det, i underlaget va det kanske fler tumörer som hade identifierats ett visst datum och sen så hade algoritmen gått på det datumet för att diagnostisera när den fick annat underlag. Så relationen eller vad man ska säga då, eller orsakssamband va ett datum men jag kan inte vad det skulle motsvara, eller jag kan inte ta ett exempel vad det skulle motsvara i elnätssammanhang men jag tänker att till exempel en sån sak är ju viktig att förstå att det inte... är ett sånt samband! Utan att, ja så det tror jag är viktigt att veta.

**79. P5:** Jag håller med [P6] och om vi också säger att energiförsörjning är någonting som liksom måste finnas där, det måste fungera. Det är liksom en grundfunktion i samhället, och sen är det väldigt dyrt om det inte fungerar för den leverantören så att dem kommer definitivt vilja veta väldigt mycket om just detta. Sen kan det hända att man kan ha det så att man kan ju ha det på olika detaljeringsnivå men det är definitivt väldigt viktigt.

**80. A: Alltså den här frågan är lite klurig tycker jag för att det är flera som vi har intervjuat som har svarat ungefär att det är viktigt att orsaken till besluten är korrekta eller adekvata men den här frågan som vi har, den här termen eller begreppet som vi har hittat i litteraturen, där handlar det om korrelationer som systemet upptäcker i omvärlden, inte korrelationer i själva systemet eller orsak till systemets beslut. Utan orsakssamband, i den data, i verkligheten, via den datan som den får in och att den hittar samband i den datan. Så om man ställer frågan på det sättet, är det relevant att hitta orsakssamband i datan?**

**81. P6:** Mm, men för om man tar det exemplet som, till exempel, med hjärntumören då, och att vad ska man ska, att algoritmen då diagnosticerar hjärntumörer på ett datum när röntgen va, då va ju det... Det är ju inte kanske tillräckligt att diagnosticera en tumör, eller det är ju inte det haha. Nu som sagt så vet jag inte något exempel på elnät, men jag tror att det kan vara liksom viktigt om man, ah att förstå. Om det skulle vara ett datum där det var särskilt många störningar och sen så tar algoritmen att när det sker det här datumet så är det, ah jag vet inte. Nä men jag tror det kan vara viktigt att...

**82. A: Det stämmer ju att de har systemen kan begå misstag.**

**83. P5:** Tänkte ni när ni pratade om information i omvärlden är det meningen då att det ska displayas i en driftcentral till exempel, flera olika informationsunderlag för beslutsfattande så att säga och om omvärldsinformation är relevant eller liksom, eller hur tänkte ni med det här? Eller är det så att systemet automatiskt fattar beslut som påverkas utav omvärldsinformation eller?

**84. A: Jag tyckte den här frågan var svår själv, Jesper kan du ge något exempel för att förtydliga den.**

**85. J: På kausalitet just eller på det här med... för att när det gäller liksom beslut så liksom vi försöker se lite generellt på smarta elnät överhuvudtaget men troligtvis kommer det behöva vara ett system där man får underlag för att fatta beslut snarare än att det fattar beslut själv, det är lite något som vi försöker se här det kanske kommer bli en del av vår diskussion i uppsatsen i fall det finns någon möjlighet att det fattar beslut själv eller om det måste vara att man ger rekommendationer. Men vad gäller kausalitet, alltså, vad det handlar om är ju att om du har en riktig korrelation som faktiskt stämmer men du vet inte vad som är orsak och vad som är verkan, så att liksom systemet har hittat någon korrelation och då... Ett exempel om man ska ta något inte kopplat till smarta elnät är liksom att när du hörs ljud så tappar jag min penna och då vet man inte, tappar jag min penna för att jag hör ett ljud eller gör pennan ett ljud när den träffar marken och sådana samband kan man se alltså den orsak-verkan är lättare att förstå om man förstår själva systemet. Förstår ni hur jag menar?**

**86. P5:** Jag tror vi förstår att det är en ganska komplex fråga faktiskt.

**87. P6:** Men jag tror att det kan, eller det är väll viktigt att förstå för att kunna bekräfta om det är vad ska man säga, rätt orsakssamband så att inte algoritmen hittar något annat som man kan säga, som man kan förstå att det inte stämmer liksom.

**88. P5:** För eran uppsats, jag tror att det är bra att, ni kanske inte ska göra det i uppsatsen men fundera kring sammanhanget som ni ställer de här frågorna i för det är olika svar beroende på, för att jag tror visst att det kan finnas beslut som kan vara väldigt viktiga och kan vara fördelaktiga att fattas automatiskt av systemet och det är när det blir avbrott av olika slag som kan kosta mycket pengar och om man då har någon slags, till exempel, det är ett exempel bara, nu spinner jag loss ganska fritt här, jag sitter inte och styr näten men jag kan tänka mig att om det är avbrott till exempel så pratar man just om att man skulle kunna ha någon slags work-around så att säga så att man tillfälligt skulle kunna, man inte behöver stänga ner hela nätet utan kanske bara en del utav det så att skademinimering så att säga, den typen av. Och det kan vara väldigt intressant med den typen av.. Och det kan behöva göras väldigt snabbt så att säga, det skulle kunna vara en sån, men någonstans är det ju så att då måste man ha ett väldigt stort förtroende för att det här görs på ett bra sätt och att det inte går att göra på ett annat sätt så att säga för att vi hinner inte. Jag tror att det finns olika, man måste nästan titta på det här i ett sammanhang.

**89. A: Du menar man svarar på frågorna för en viss applikation**

**90. P5:** Ja, för en viss applikation i ett visst sammanhang så att säga.

**91. A: Eller ett visst användningsområde, liksom.**

**92. P5:** Pratar du om att styra nätet och ta upp information så kan omvärldsinformation som väder vara väldigt relevant till exempel eller att det händer saker och ting i omvärlden som gör att du kanske måste säga att ett kärnkraftverk lägger ner, eller säg att vinden inte blåser då kanske det är väldigt relevant information för att styra ett nät. Så att jag menar, kanske lite mer generella exempel men den typen av information är superviktig. Sen kanske det inte är så intressant vad som orsakar det hela men verkan är ju liksom ganska... har stor betydelse.

**93. J: Ja, ska vi gå vidare på nästa? Då är det överförbarhet, som innebär att man kan använda lärdomar man... eller alltså om systemet har lärt sig att lösa ett problem så kan man använda det för att lösa liknande problem. Är det något ni tror viktigt i smarta elnät?**

**94. A: Och då kan ni svara för, om ni vill, för ett särskilt användningsområde om ni kommer och tänka på något särskilt användningsområde. Om det underlättar för att svara.**

**95. P5:** Alltså själva grunden med intelligens är ju att det är intelligens och att det ska kunna hitta lösningar och ja, frågan är hur mycket man vill släppa systemet loss så att säga. Det är ju viktigt, men tillförlitlighet är viktigare i min kalender så att säga.

**96. P6:** Ja precis, jag har nog svårt att svara eller spontant har jag nog lite svårt att svara på den frågan men i relation till dem andra så skulle jag tro att det kanske inte är.... lika viktigt. Men den var svår att svara på tycker jag.

**97. P5:** Jag håller med dig [P6].

**98. J: Jo dem är lite svåra ibland men, ja...**

**99. A: Vi har ju inte sållat på förhand vilka frågor som är relevanta för smarta elnät för att vi ville inte göra antaganden. Så därför kan det komma upp frågor som verkar inte vara relevanta.**

**100. P6:** Nä eller alltså jag tror absolut att det är en relevant...

**101. P5:** Relevant är det definitivt.

**102. P6:** Absolut men jag skulle tro att det...

**103. A: Inte lika viktigt.**

**104. P6:** Ah, precis det skulle jag kanske dra en slutsats [inaudiable].

**105. P5:** Nä men och kan systemet spotta ut tre lösningar som har fungerat tidigare i andra sammanhang på det här problemet så är det väll suveränt till exempel för då överför ju den kunskap som den har sedan tidigare och dra slutsatsen att den kan använda samma lösningar igen så att säga för att det är liknande parametrar och sen kanske ställer frågor "men hur gör vi med det här för att här finns det någon missmatch mot tidigare case så att säga. Det är väll jättebra. Men

om du frågar efter om man ska vikta det hela så är ju tillförlitlighet och leverans av el det är ju liksom A och O.

**106. J: Ja, då tänker jag att vi får vidare till informativitet som innebär att man får information om hur ett beslut har fattats så att man kan bland annat granska om det är korrekt och man kan liksom dra lärdomar av hur själva beslutet har fattats. Ett exempel här är liksom om en läkare ser, om man tar det här tidigare med scanningsbilderna, så vill man ju veta varför tror man att den här personen har cancer, man vill inte bara veta "den här patienten har cancer" för man kan lära sig att upptäcka cancer och dels för att man kan se är det att den har gått bara på datum eller är det liksom, ja.**

**107. P5:** Det är väll jätteviktigt med spårbarhet men frågan är, man kanske inte behöver det just i beslutsituationen utan det kanske är så att man behöver spårbarheten efteråt när man har fattat besluten och kunna följa upp, dokumentation eller likande. Så att frågan är hur mycket information klarar man av att hantera om det är beslutsfattaren som ska styra nätet eller om det är spårbarhet efteråt, varför gick det fel, så att säga. Eller om den här spårbarheten gör att man kan, inte vet jag, följa upp någonting så att säga, då kanske man vill göra det liksom i vissa fall men i andra fall... det kanske ska ligga längre ner i systemet om jag säger så, men det är ju viktigt att det finns där.

**108. P6:** Ja, jag tror att det kan vara, nu känns det som man säger viktigt på allting men jag tror att det kan vara viktigt. Alternativt att man får, men så klart att det också igen då beror på sammanhanget, alltså inom vilket område. Till exempel om man ska ta liksom ett beslut i kontrollrummet där man sitter och driftar nätet. För hur man ska koppla och så vidare, men om man också har annan input, alltså säg att man har en algoritm för att rekommendera ett beslut "gör såhär", och så får man inte veta varför, eller en analys men att man också då har annan data, då kanske man kan acceptera att man inte vet. Men jag tror ändå att det kan vara viktigt. Också för elnät, alltså för att kunna förstå. Oftast behöver man ju förklara på något sätt eller... och dem besluten man fattar beror ju kanske också lite på så jag tror att det kan vara viktigt att också få veta varför.

**109. P5:** Det handlar precis som [P6] säger om varför men sen också det här med att man ska kunna återställa om någonting är fel också. Jag menar, ta aktiemarknaden, om den dyker så måste man ju återställa den på den nivån den var tidigare på något sätt. Det är likadant med nätet, om någonting inträffar så måste du kunna återställa felet eller ha någon slags dokumentation på det hela och det kanske var ett beslut någonstans men det måste ju någonstans ha en dokumentation.

**110. A: Kan man tänka sig att det här, jag tänker mig att det här borde kunna relateras till tillit. Att ju mer man litar på systemet desto mindre är betydelsen av att förstå precis hur den kom fram till beslutet.**

**111. P5:** Ja, jag håller med på det. [inaudible] Vilket man kan göra de första fem gångerna och sen vet man att det fungerar bara.

**112. J:** Kan vi gå vidare?

**113. P5:** Absolut jag har tyvärr bara tre minuter eller två sen måste jag ta nästa möte

**114. J: Då får vi skynda oss, då har vi rättvist och etiskt beslutsfattande. Ni kanske vet att det blir biaser i AI för att man har data som reflekterar biaser och sånt och den kan också fatta oetiska beslut, marknadsföringsbotar kan marknadsföra alkohol till alkoholister, tror**

**ni det är någonting som är relevant här. Kan det vara något som är viktigt att stoppa för att det kan råka komma annars.**

**115. P6:** Jag tycker ju att det är viktigt men jag har lite svårt att se, hmm. Jag vet inte riktigt hur man kan tänka sig exemplet inom elnät men det känns som en viktigt grej generellt med AI att vara uppmärksam på.

**116. P5:** Jag håller med.

**117. J: Det är framförallt när man riktar sig till privatpersoner så det kanske inte händer så mycket här liksom. Sen så har vi tillgänglighet som handlar om att folk som inte är experter eller är särskilt insatta i systemet kan förstå det. Kanske tillexempel, om ni vill granska företagssystem.**

**118. P6:** Ja men kanske om man kopplar det lite till kompetens... eller jo men jag tror det kan vara viktigt om man ser liksom en större implementering av det och att alla kanske inte har... Inom elnät är det ju så brett liksom vad man sitter och jobbar med, eller den egna kompetens eller kunskapsområdet och om då artificial intelligens kan vara som ett stöd och verktyg till det så tror jag att det är viktigt för att man inte ska ha behöva ha studerat artificiell intelligens i fem år, alltså att man ändå kan förstå.

**119. P5:** Jag tänkte på oetiska beslut bara, det var integriteten som du var inne på [P6] det kommer ju in där någonstans och hur man använder data.

**120. P6:** Ja jag funderar på, det här är bara en fundering och en spekulation men tillexempel om det skulle komma in, jag vet inte om eller hur det skulle komma, men man får ju, nu ska vi se här. Hur man väljer att bygga ut elnät tillexempel, det ska ju vara, det är ju ett demokratiskt system så det ska ju liksom finnas för alla och man ska ju liksom inte välja att bygga ut något till något område för att det finns, att det är finare på något sätt, alltså socioekonomiskt till exempel eller så där.

**121. A: Så att dem är större förbrukare av el tillexempel kanske?**

**122. P6:** Det är väll liksom mer liksom ett behov av, vart det finns behov av förnyelse av elnätet ur ett tekniskt perspektiv. Och också så klart när man tänker så här, då har man mycket dialog med staten... man får liksom inte prioritera något, socialgrupp eller socioekonomisk område över något annat om det skulle... jag har ingen aning, det här är ju som sagt bara en spekulation men om det på något sätt skulle komma in i prognostisering av elnät då är ju det... just med bias och sånt där att det är viktigt att det inte...

**123. P5:** Det gör ju det [P6] om man tänker på att det kostar en mängd, om vi måste investera i mer sensorer på konsumentnivå eller att vissa områden har mer råd till att... ladda bilen eller vad vet jag, investera i infrastruktur för att kunna ha, ladda bilen eller någonting eller köpa vissa typer av lösningar eller för att få en billigare konsumtionskostnad eller någonting sånt. Jag måste tyvärr gå ur mötet och gå in i nästa men [P6] har du några möjligheter att sitta kvar några minuter, jag vet inte hur det ser ut för dig.

**124. P6:** Jo men jag kan sitta kvar några minuter till.

**125. P5:** Alexander och Jesper tack så mycket för att ni hörde av er.

**126. J:** Ja, tack själv.

[...]

**127. J: Vi pratade om tillgänglighet, du var klar där va?**

**128. P6:** Mm

**129. J: Och så har vi interaktivitet, som innebär att det blir lättare för folk att interagera med systemet, till exempel kunder, eller samarbetspartners. Tror du det är viktigt?**

**130. P6:** Ja men absolut, där skulle det ju kunna vara... Jag tror att det skulle kunna finnas... Om man tänker sig också att man skulle typ ha det som ett stöd i ett driftsammanhang då kanske det skulle det kunna vara viktigt att interagera med systemet. Och sen kanske, för i vissa, man har ju vissa angreppspunkter mellan elnätsbolag där jag funderar på om det skulle kunna vara... men jag vet inte, hur viktigt det är mot...[inaudiable] jag kan se möjligheter med det och så vet jag inte ett förstå steg hur viktigt man kanske tycker att det är. Men absolut om man tänker sig då att det är... jo men kanske, ah men viktigt i vissa sammanhang.

**131. A: Inte lika viktigt som några av de andra punkterna, eller?**

**132. P6:** Ja men kanske lite mindre viktigt.

**133. J: Då, det sista är integritetsmedvetenhet som du har varit inne på. Det är ju det här att liksom, AI:t kan ju liksom hitta samband som råkar kränka folks integritet utan att man, just med deep learning så kan det ju hända utan att man vet det. Har du något mer som du vill säga där.**

**134. P6:** Är det om det är viktigt? Ja precis och då tänker jag väll amen precis som jag var inne på tidigare på framförallt kopplat till smarta mätare, eller mätare, att där kommer en aspekt av integritet in och där är det ju så klart väldigt viktigt.

**135. J: Det var våra frågor, tack så mycket. Har du något du vill tillägga eller så?**

**136. P6:** Ja eller jag kan… lite generellt så tror jag att det är ett, eller jag uppfattar inom för elnätsbolag som ett ganska nytt område men som vi var inne på att man absolut ser potentialen i det och om man tänker såhär också generellt inom elnätsområdet så har ju det förändrats väldigt mycket och väldigt mycket det senaste egentligen och det blir lite mer komplex branch så jag tror att man absolut ser stora nyttor med att tillämpa den här tekniken och man ser ju det också att det är liksom... men typ som de exemplen som vi tog i början det är någonting som också är väldigt nytt och en del var exempel på forskningsprojekt där man tittar på det och en del är att det finns vissa, till exempel inom förvaltning och inom prognostisering att det finns liksom vissa exempel i verkligheten eller vad man ska säga också kring det. Men väldigt spännande område.

# References

Albadi, M.H., & El-Saadany, E.F. (2018). A summary of demand response in electricity markets, *Electric power systems research,* vol. 78, no. 11, pp. 1989-1996, Available online: https://www.sciencedirect.com/science/article/pii/S0378779608001272 [Accessed 11 May 2020]

Anwar, T., Sharma, B., Chakraborty, K., & Sirohia, H. (2018). Introduction to Load Forecasting, *International Journal of Pure and Applied Mathematics,* vol. 119, no. 15, pp. 1527-1538, Available online: https://acadpubl.eu/hub/2018-119-15/3/567.pdf [Accessed 11 May 2020]

Arbetsmiljöverket. (2020). Arbete vid högspänningsledningar, Available online: https://www.av.se/produktion-industri-och-logistik/bygg/risker-vid-byggnad--och anlaggningsarbeten/arbetsmiljoplan-och-dess-risker/arbete-vid-hogspanningsledningar/ [Accessed 14 May 2020]

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, *Information Fusion,* vol. 58, pp. 82-115, Available online: https://www.sciencedirect.com/science/article/pii/S1566253519308103 [Accessed: 11 May 2020]

Brynjolfsson, E., & McAfee, A. (2014). The Second Machine Age: Work, progress, and prosperity in a time of brilliant technologies, USA: WW Norton & Company

Bunge, M. (1963). A general black-box theory, *Philosophy of Science,* vol. 30, no. 4, pp. 346-358, Available online: https://www.jstor.org/stable/186066 [Accessed 11 May 2020]

Chakraborty, S., Tomsett, R., Raghavendra, R., Harborne, D., Alzantot, M., Cerutti, F., Srivastava, M., Preece, A., Julier, S., Rao, R. M., Kelley, T. D., Braines, D., Sensoy, M., Willis, C. J., & Gurram, P. (2017). Interpretability of deep learning models: A survey of results, *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI),* pp. 1-6, Available online: https://ieeexplore-ieee-org.ludwig.lub.lu.se/document/8397411 [Accessed 14 May 2020]

Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., & Nel, P. (2018). Notes From the AI Frontier: Insights from hundreds of use cases, USA: McKinsey & Company, Available online: https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20AI%20frontier%20Applications%20and%20valu

e%20of%20deep%20learning/Notes-from-the-AI-frontier-Insights-from-hundreds-of-use-cases-Discussion-paper.ashx [Accessed 2 April 2020]

Ellevio. (n.d.). Ellevio i korthet, Available online: https://www.ellevio.se/om-oss/ellevio-i-korthet/ [Accessed 16 May 2020]

Energimyndigheten. (2009). Funktionskrav inom elförsörjningen, Available online: https://energimyndigheten.a-w2m.se/FolderContents.mvc/Download?ResourceId=2405 [Accessed 14 May 2020]

Energimyndigheten. (2020). Hållbar energi för alla, Available online: http://www.energimyndigheten.se/om-oss/ [Accessed 16 May 2020]

Evans, R & Gao, J. (2016). DeepMind AI Reduces Google Data Centre Cooling Bill by 40%, web blog post, Available at: https://deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill-40 [Accessed 11 May 2020]

Fang, X., Misra, S., Xue, G., & Yang, D. (2012). Smart Grid — The New and Improved Power Grid: A Survey, *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 944-980, Available online: https://ieeexplore-ieee-org.ludwig.lub.lu.se/abstract/document/6099519 [Accessed 11 May 2020]

Harbers, M.,van den Bosch, K., & Meyer, J. (2010). Design and Evaluation of Explainable BDI Agents, *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pp. 125-132, Available online: https://ieeexplore-ieee-org.ludwig.lub.lu.se/document/5614190 [Accessed 19 May 2020]

He, Y., Mendis, G. J., & Wei, J. (2017). Real-Time Detection of False Data Injection Attacks in Smart Grid: A Deep Learning-Based Intelligent Mechanism, *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505-2516, Available online: https://ieeexplore.ieee.org/abstract/document/7926429 [Accessed 11 May 2020]

IEA. (2010). Renewable Energy Essentials: Hydropower, Available online: https://web.archive.org/web/20170329132409/http://www.iea.org/publications/freepublications/publication/hydropower_essentials.pdf [Accessed 2 April 2020]

IEA. (2019). Renewables, Available online: https://www.iea.org/fuels-and-technologies/renewables [Accessed 11 May 2020]

Jindal, A., Aujla, G. S., Kumar, N., Prodan R., & Obaidat, M. S. (2018). DRUMS: Demand Response Management in a Smart City Using Deep Learning and SVR, *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6, Available online: https://ieeexplore-ieee-org.ludwig.lub.lu.se/document/8647926 [Accessed 11 May 2020]

Langley, P., Meadows, B., Sridharan, M., & Choi, D. (2017). Explainable Agency for Intelligent Autonomous Systems, *Innovative Applications of Artificial Intelligence Twenty-Ninth IAAI Conference*, Available online: https://aaai.org/ocs/index.php/IAAI/IAAI17/paper/view/15046 [Accessed 19 May 2020]

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning, *Nature*, vol. 521, no. 7553, pp. 436–444, Available online: https://doi.org/10.1038/nature14539 [Accessed 11 May 2020]

Lipton, Z. C. (2016). The mythos of model interpretability, *Queue*, vol. 16, no. 3, pp. 31-57, Available online: https://arxiv.org/abs/1606.03490 [Accessed 11 May 2020]

Liu, Y., Ning, P., & Reiter, M. K. (2011). False data injection attacks against state estimation in electric power grids, *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, pp. 1-33, Available online: https://dl.acm.org/doi/abs/10.1145/1952982.1952995 [Accessed 11 May 2020]

Lu, R. & Hong, S, H. (2019). Incentive-based demand response for smart grid with reinforcement learning and deep neural network, *Applied Energy*, vol. 236, pp. 937–949, Available online: https://www.sciencedirect.com/science/article/pii/S0306261918318798 [Accessed 11 May 2020]

Lin, J., Yu, W., Yang, X., Xu, G., & Zhao, W. (2012). On False Data Injection Attacks against Distributed Energy Routing in Smart Grid, *IEEE/ACM Third International Conference on Cyber-Physical Systems*, pp. 183-192, Available online: https://ieeexplore.ieee.org/abstract/document/6197400 [Accessed 11 May 2020]

Miljödepartementet. (2007). Förhandsprövning av nättariffer m.m., SOU:2007:99, Available online: https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2007/12/sou200799/ [Accessed 16 May 2020]

Mälarenergi. (n.d.). Mälarenergi - 150 år i Västeråsarnas tjänst, Available online: https://www.malarenergi.se/om-malarenergi/malarenergi/ [Accessed 16 May 2020]

Nielsen, M. A. (2015). "Neural Networks and Deep Learning", San Francisco: Determination Press, Available online: http://neuralnetworksanddeeplearning.com/ [Accessed 18 May 2020]

NIST. (2010). NIST Framework and Roadmap for Smart Grid Interoperability Standards. Available online: https://www.nist.gov/system/files/documents/public_affairs/releases/smartgrid_interoperability_final.pdf [Accessed 2 April 2020]

Niu, X., Li J., Sun J. & Tomsovic, K. (2019). Dynamic Detection of False Data Injection Attack in Smart Grid using Deep Learning, *2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1-6, Available online: https://ieeexplore.ieee.org/abstract/document/8791598 [Accessed 11 May 2020]

Oates, B. J. (2006). Researching information systems and computing, Middlesborough: Sage.

Odena, A., Goodfellow, I. (2018). TensorFuzz: Debugging neural networks with coverage-guided fuzzing, *arXiv preprint*, Available online: https://arxiv.org/abs/1807.10875 [Accessed 14 April 2020]

Regeringen. (2016). Ramöverenskommelse mellan Socialdemokraterna, Moderaterna, Miljöpartiet de gröna, Centerpartiet och Kristdemokraterna, Available online: https://www.regeringen.se/contentassets/b88f0d28eb0e48e39eb4411de2aabe76/energi overenskommelse-20160610.pdf [Accessed 11 May 2020]

Ruelens, F., Claessens, B. J., Vrancx, P., Spiessens, F., & Deconinck, G. (2019). Direct load control of thermostatically controlled loads based on sparse observations using deep reinforcement learning, *CSEE Journal of Power and Energy Systems*, vol. 5, no. 4, pp. 423-432, Available online: https://ieeexplore.ieee.org/abstract/document/8928284 [Accessed 11 May 2020]

Sanderson, G. (2017a) What is backpropagation really doing? | Deep learning, chapter 3 [Video online], Available at: https://youtu.be/Ilg3gGewQ5U [Accessed 14 May 2020]

Sanderson, G. (2017b) Gradient descent, how neural networks learn | Deep learning, chapter 2 | Deep learning, chapter 3 [Video online], Available at: https://youtu.be/IHZwWFHWa-w [Accessed 14 May 2020]

Shi, H., Xu, M., & Li, R. (2017). Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN, *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271-5280, Available online: https://ieeexplore-ieee-org.ludwig.lub.lu.se/document/7885096 [Accessed: 11 May 2020]

Shwartz-Ziv, R., & Tishby, N. (2017). Opening the black box of deep neural networks via information, *arXiv preprint*, Available online: https://arxiv.org/abs/1703.00810 [Accessed 11 May 2020]

Sitawarin, C., Bhagoji, A. N., Mosenia, A., Chiang, M. & Mittal, P. (2018). DARTS: Deceiving Autonomous Cars with Toxic Signs, *arXiv preprint*, Available online: https://arxiv.org/pdf/1802.06430.pdf [Accessed 14 April 2020]

Sun, Y., Huang, X., Kroening, D., Sharp, J., Hill, M. & Ashmore, R. (2018). Testing Deep Neural Networks, *arXiv preprint*, Available online: https://arxiv.org/pdf/1803.04792.pdf [Accessed 11 May 2020]

Sun, Y., Huang, X., Kroening, D., Sharp, J., Hill, M. and Ashmore, R. (2019). DeepConcolic: Testing and Debugging Deep Neural Networks, *2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, pp. 111-114, Available online: https://ieeexplore.ieee.org/abstract/document/8802786 [Accessed 14 April 2020]

Tuballa, M. L., & Abundo, M. L. (2016). A review of the development of Smart Grid technologies. *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 710-725, Available online: https://www.sciencedirect.com/science/article/pii/S1364032116000393 [Accessed 18 May 2020]

Vattenfall Eldistribution. (n.d.). Om Vattenfall Eldistribution, Available online: https://www.vattenfalleldistribution.se/om-oss/ [Accessed 16 May 2020]

Zhang, D., Han, X. & Deng, C. (2018). Review on the research and practice of deep learning and reinforcement learning in smart grids, *CSEE Journal of Power and Energy Systems*, vol. 4, no. 3, pp. 362-370, Available online: https://ieeexplore.ieee.org/abstract/document/8468674 [Accessed 11 May 2020]

Zhao, X., Banks, A., Sharp, J., Robu, V., Flynn, D., Fisher, M., Huang, X. (2020). A Safety Framework for Critical Systems Utilising Deep Neural Networks, *arXiv preprint*, Available online: https://arxiv.org/abs/2003.05311 [Accessed 14 april]