Ethical Considerations for Implementing AI-Driven Building Automation Systems

Course No: LE1-013
Credit: 1 PDH

Joseph Garon, P.E., HFDP

engineering.com
PDH FOR THE PROFESSIONAL

Continuing Education and Development, Inc.

P: (877) 322-5800
info@cedengineering.com

Introduction

The advent of digital controls allowed for building systems to be fully automated allowing for improved operation, maintenance, and efficiency. Now, Artificial intelligence (AI) is transforming the realm of building automation systems. The demand for AI-driven building automation systems (BAS) has increased as regulatory pressure continues to emphasize sustainability and green building practices. AI is likely to become a critical tool for optimizing building performance such as enhancing occupant comfort, improving energy efficiency, and reducing operating costs. This is due to AI's potential in creating intelligent environments that can adapt to real-time conditions. However, while this technology has many potential upsides and applications, there are inherent risks and ethical considerations. This course will discuss the ethical considerations of AI as it specially relates to building systems. It will also serve as an introduction to how AI systems function and their capabilities.

What Is AI And How Does It Work

AI is fundamentally a series of algorithms that are designed to mimic human cognitive functions. By using data and high-level computation power, AI can solve complex problems, recognize patterns, and make decisions. The following figure provides an overview for the AI model development and training which will be further explained in the next sections:

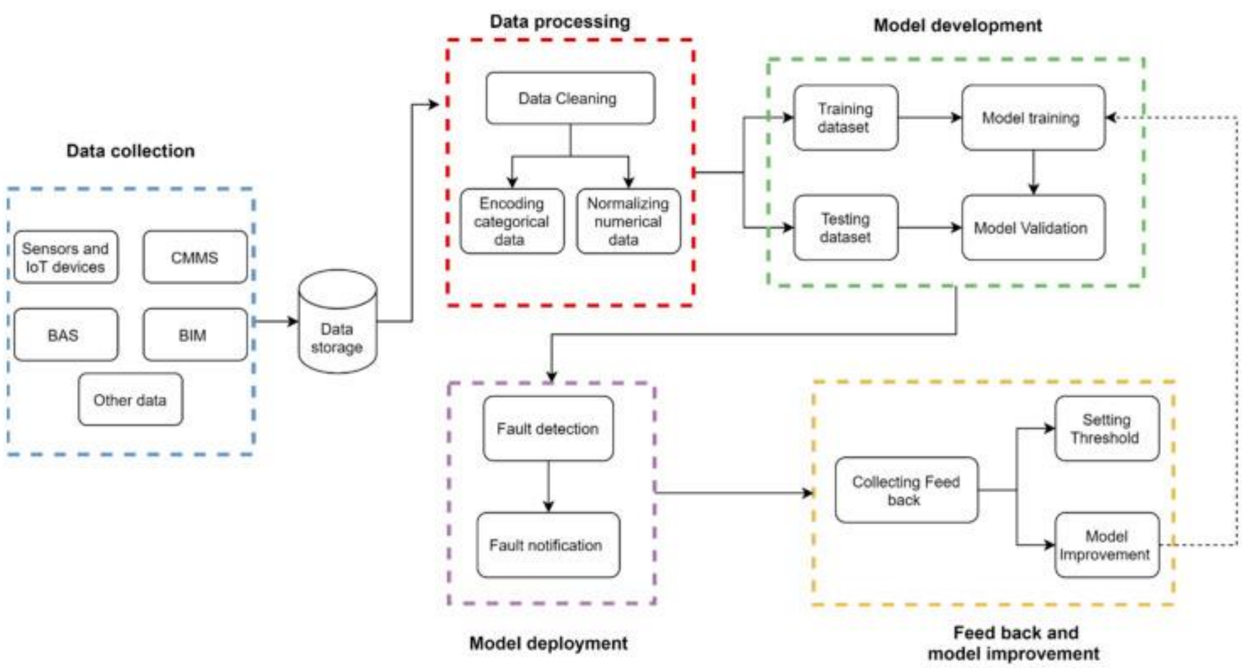

Figure 1: AI Model Architecture and Framework (source: Bouabdallaoui, et. All)

Data Collection and Preprocessing

An AI system initially requires the collection of substantial amounts of data. The quality, quantity, and diversity of data will directly influence the performance of an AI model. In other words, the volume of data and the relevance will determine how well the AI system is likely to perform. In the context of BAS, this could be from control devices in the field such as actuators, thermostats, occupancy sensors, etc. As more data is obtained, the more accurate the system is likely to become. Conversely, the larger the volume of data, the more challenging it is to store.

Building systems are complex since they are influenced by human activity and so day-to-day operation can be highly diversified. In addition, to collecting large quantities of data, the variety of data is also important to give the AI context. Below is a table the provides some categories and examples of data variety that may be found in an AI-driven building automation system.

Table 1: Examples and Applications of Various Data Used By AI-Driven Building Systems

Data Type	**Examples**	**Application**
Environmental	Temperature, Humidity, C02 or Other Contaminant Levels, Light levels, Pressurization	Adjusting HVAC systems to maintain optimum indoor air quality and comfort levels.
Occupancy	Access Control, Occupancy Sensors, Motion Sensors	Adjust setting based on occupant load throughout various areas of a building
Energy Consumption	Utility Data and Energy Meters	Balancing supply and demand to reduce building energy costs
Security	Access Control, Security Cameras, Alarms	Ensuring building security by responding to unauthorized access or threats
Maintenance	Fault Detection, Equipment Status	Predictive maintenance
Weather	Temperature, Humidity, Barometric Pressure, Air Quality, Wind Speed, Solar Radiation	Predictive adjustments to reduce energy consumption
User Preference	Temperature, Humidity, Lighting, and Ventilation Setpoints	Adjusting system to maintain occupant comfort
Building Design	Building Plans (Architectural, MEP, Structural, etc.) and Building Standards (Insulation, Piping Materials, etc.)	Optimize energy use and suggest retrofits or upgrades.

Historical	Energy Consumption, Maintenance, and Occupancy Trends	Builds AI model accuracy to better predict future operation.
External	Utility Prices and Incentives	Adjust energy usage based on pricing and demand for grid stability.

High quality data is important to provide accuracy as well as prevent biased or incorrect models. Some examples of data quality may take the form of accuracy, completeness, timeliness, consistency, and relevance. Accuracy and completeness are self-explanatory. Of course, if the data compiled is not accurate or complete the AI will create inherent biases. Timeliness refers to providing real-time data. Delays in data due to latency could create inefficiencies such over conditioning a part of the building that is unoccupied. Consistency refers to delivering the data in the same or similar format throughout the system. This allows for seamless integration to prevent errors in AI decision making. Finally, data that is irrelevant to system operation will clutter the AI processing and potentially reduce efficiency.

All this raw data that is collected is rarely ready for immediate use in AI models. It is necessary to preprocess the data. This means that the data is cleaned, transformed, and prepared in a usable format so the AI model can learn from it effectively. The preprocessing step addresses the issues of missing data, noise, and inconsistencies which would degrade performance.

Data cleaning refers to the handling of missing data and the removal of noise. Human error and sensor failures may result in missing values. In addition, naturally occurring feedback loops in the system or inherent noise in the measurement devices can create fluctuations in the data. Missing data can be dealt with by removing incomplete records or using imputation. **Imputation** is a technique which replaces missing values with a mean, median, or predicted value. Some data sets might contain irrelevant or erroneous information which is referred to as **noise**. Noise can be reduced by filtering out outliers or correcting erroneous entries. Another noise reduction technique is **data smoothing** which refers to the elimination of short-term fluctuations or anomalies within a data set. There are various methods to perform data smoothing. The specifics of these techniques are beyond the scope of this course, but these are basically different techniques to apply curve fits to trendlines to remove fluctuations. The below figure demonstrates an example of data smoothing applied to global temperature (compared to a long-term average) and the year recorded. The green line is the “smoothed” curve which would be used to predict future numbers.

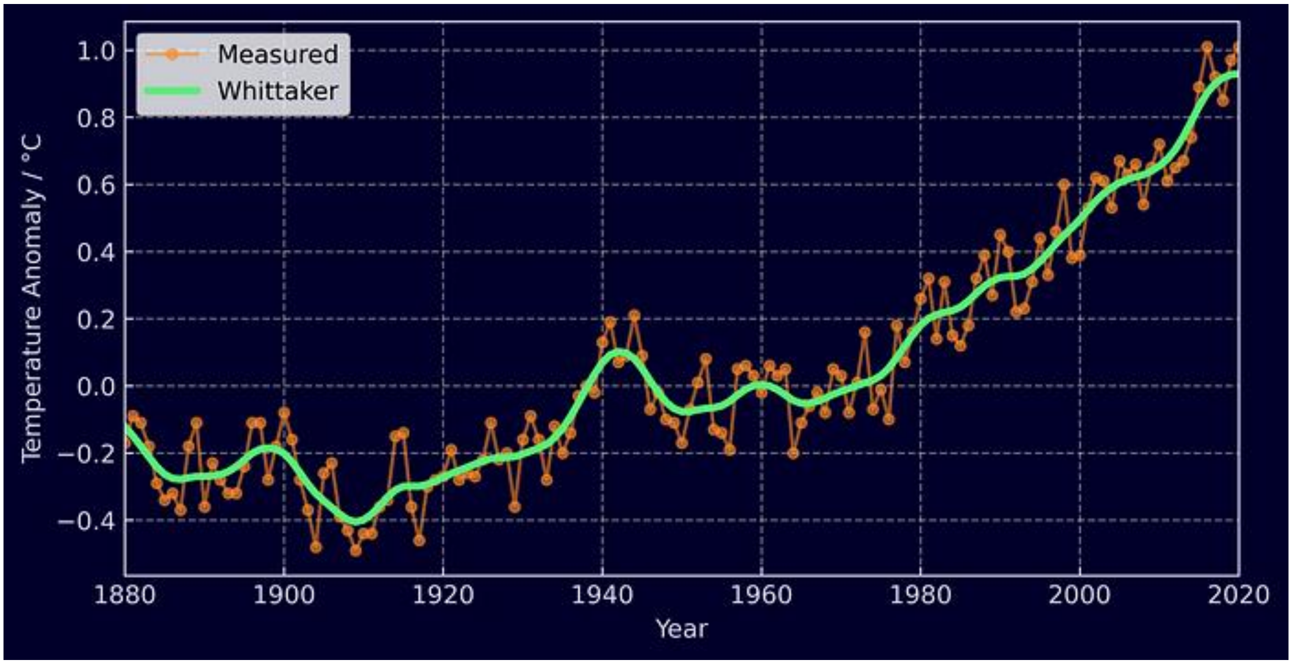

Figure 2: Example of Data Smoothing (Bowell)

Next, is data transformation. Data transformation can involves normalization and encoding categorical data. **Normalization** scales data within a specific range or distribution to ensure that all features (data elements) contribute to the model equally. For example, one data set may have a range of numbers in the thousands while another data set may have numerical values in single digits. The normalization function ensures that the model will not disproportionately focus on the larger values. Encoding converts data to a specific format so that it can be more efficiently used by the AI. Data can come in various forms besides numbers. Categorical data for example is data that is understood with text such as "air handling unit," "VAV," "cooling," or "heating." A computer cannot understand text as humans can, therefore, this data must be converted into a numerical format so that the AI can use it in the algorithm.

Algorithms and Models

As mentioned in the intro, the "brain" of the AI is the algorithm. It is the algorithms that allow the AI to learn from the collected data. This is referred to as **machine learning**. There are different types of machine learning and levels of AI sophistication. While the specifics of machine learning are beyond the scope of this course, a brief overview of general categories is discussed below.

Broadly speaking AI algorithms can be categorized as supervised or unsupervised leaning. These forms of learning are usually performed using fixed data sets. **Supervised learning** is where the AI is trained on labeled data to predict outcomes and classify data points. In other words, the model is trained to associate an input with an output. This must require a human "supervisor" to label or classify the data. **Unsupervised learning** refers to when the AI finds patterns and relationships in unlabeled data. The AI then learns useful properties derived from that dataset using probability distributions without human input.

R**einforcement learning** is where AI learns through a trial-and-error approach. The AI will receive rewards or penalties based on its actions. These types of algorithms do not experience a fixed data set and are intended to be reactive to their environment.

An AI-driven system will likely use a combination of these learning algorithms depending on the degree of complexity that is employed.

Training the model

As already explained, AI is trained by feeding it data and adjusting parameters to minimize predictive errors or classifications. This is typically done by using **gradient descent** which is a technique that involves the model iteratively reducing the difference between its predictions and actual outcomes.

Once the AI model is trained on the data , the model is validated using separate data that was not used during the model training. This validation step is meant to assess the model's performance. This test also ensures that the model can generalize and adapt to data it has not seen.

After training, the model can make predictions and/or decisions based on new input data. The AI will continuously learn as more data is processed. This continuous learning further improves accuracy and performance. The trained model is integrated into the BAS where it can begin to automate tasks. The continuous learning of the AI is monitored and fine-tuned by user feedback and more data. This ensures that the model will remain relevant and effective.

The Capabilities of AI-Driven Building Systems

Within the context of building systems, AI functions as a predictive tool that anticipates changes based on the data it accumulates. The more data that is fed into the system, the better the AI can predict and make real-time changes to the system. Therefore, the AI becomes better the longer it is used. Usually, these changes occur as micro-adjustments to setpoints. By making these small adjustments continuously over an extended period of time, the building performance is optimized to a higher level than what manual inputs from a human could achieve.

For most buildings, much of the energy is used to power the HVAC system. Traditional HVAC systems will operate on a fixed schedule or will respond to basic inputs from control devices in the field such as maintaining a set point on a discharge air temperature sensor. An AI-driven system can produce a vast amount of data using these same control devices while incorporating data on occupancy patterns, weather forecasts, and energy prices. The result is a reduction in energy costs without sacrificing occupant comfort.

Lighting control in buildings can be optimized by adjusting light levels based on occupancy, availability of natural light, and user preferences. By learning from occupant behavior, AI systems can reduce energy waste by only using lighting when necessary. It can also contribute to occupant health by adjusting the color temperatures throughout the day to align with human circadian rhythms. The result is enhanced comfort levels for the occupant and increased productivity.

Just like with HVAC, AI can enhance energy efficiency in plumbing systems by optimizing domestic hot water systems based on usage. In addition, there is potential to use AI for leak detection by monitoring water usage patterns to identify anomalies. Not only would this lead to reduced water usage but also prevent water damage or reduce water damage.

With full integration of AI in to building automation systems (BAS), AI-driven systems can lead to proactive building management by utilizing predictive maintenance and recommending retrofits or upgrades. This means prolonging system lifespans and reducing downtime.

Ethical Considerations

Energy Optimization

One of the most touted functions of an AI-driven system is the improved energy performance for a building. Engineers and regulatory agencies participate globally to reduce carbon footprints and reduce energy consumption with the goal of reducing man-made climate change. This entails reducing reliance on energy sources that rely on fossil fuels. The main question to consider is whether AI systems actually contribute to this sustainable practice.

The **rebound effect** is the reduction of expected gains in energy efficiency from new technologies meant to increase efficiency as a result of behavioral or systemic responses. It is a widely documented phenomenon that was even predicted as far back as 1865, before the industrial revolution. The most common example of this is when a more fuel-efficient car is purchased only to result in the vehicle being driven more frequently. Examples of a rebound effect might happen at multiple levels. For instance, a direct rebound effect would be if the AI is able to optimize setpoints to reduce energy consumption and thus lower the energy bill. This may prompt building managers or occupants to set more comfortable (more energy intensive) settings.

Indirect consequences from the rebound effect would involve the goods and services used to produce the AI. AI relies heavily on data collection and computation power. This requires servers and processors which, at scale, consume exorbitant amounts of energy. So, while energy is reduced locally at the building level, the energy is consumed remotely at a data warehouse. For example, in a study performed by the University of Massachusetts Amherst, it was found that training a single AI model could emit as much CO2 as five cars during their lifetimes. To be clear, the AI models used in this study were neural network NLP (natural language processors) which were designed for high levels of accuracy. It was found that once the network achieved a certain level of accuracy, increasing the level of accuracy further created exponentially more emissions. If ethically, the engineer is concerned with climate change, the reduction in energy through an AI-driven BAS may not create net energy savings.

This paradoxical relationship raises concerns about the long-term sustainability of AI-driven systems when these systems are applied at scale. The rebound effect can also be applied on a macro-economic level. Reductions in energy costs could lead to increased economic activity which then, in turn, leads to higher energy demand (like the car example).

Consumption	**CO_2e (lbs)**
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

Figure 3: AI CO2 Consumption Comparison (Strubell)

There is also an important ethical consideration that engineers ensure that the reduction in energy use does not compromise the indoor air quality or occupant comfort for the sake of energy efficiency. While turning a system off or reducing outside air may reduce building energy consumption, there is an inherent risk in jeopardizing occupant health such as in a healthcare facility. It also may not be in the building owner's best interest to raise system temperatures if it were to lead to occupant discomfort. Increased occupant discomfort is proportionally linked to a reduction in productivity which may be more costly than the money saved on the energy bill.

<u>Data Privacy</u>

AI-driven BAS raises concerns about how data is used, stored, and shared. Data is the backbone for the AI model. The more data that is acquired, the more accurate model. This would include potential personal data of the occupant such as when individuals enter or leave a room, how long they stay in a specific area of a building, or what their personal preferences are (temperature settings, light usage, water usage). Users may ultimately not find the idea of increased surveillance appealing and consider it a violation of their privacy.

Systems should be designed to protect the user and occupant privacy so that this risk is minimized. Furthermore, users should be made fully aware of what data is collected and have some level of control over this information. This is called **informed consent**. An ethical practice for data privacy would be to make these systems transparent and provide built-in consent features so it is clear from the user what data is collected, why it's being used, and allow an occupant or user to opt-out. Of course, opting out of these features could potentially have a negative impact on energy efficiency.

Another consideration is data security. While there is always some risk of a data breach even with a standard BAS, the volume of data available for mining by a malicious actor could certainly be used for exploitative purposes. Some institutions may find data security more of a concern than others (government buildings in particular). Systems that utilize AI must consider adopting higher levels of security to help mitigate potential breaches. This could mean stronger encryption protocols, regular updates to patch software vulnerabilities, and conducting security audits.

Furthermore, some thought should be directed towards contingency plans for when data breaches occur so that action may take place quickly to minimize harm. For owners, this also means considering how the AI data is secured, whether it is via third-party proprietary network or if it is managed in-house.

Malicious actors may not be merely external threats, but they may also be those who have legitimate access to the data. Data managers or administrators may use collected data in ways that go beyond the intended purpose. For example, occupancy information could be used for monitoring employee productivity or behavior. Data might also be used for commercial gain.

A real-world example of this was Sidewalk Labs. Sidewalk Labs was a subsidiary of Alphabet Inc. and proposed a "smart city" to be built along Toronto's waterfront area. Sensors would be embedded throughout the area and monitor pedestrian movements, energy use, air pollution, waste management, and even vital signs. Ultimately, concerns about how the data was going to be used and shared (without the residents' consent) led to the projects undoing.

Engineers must maintain and uphold the public's trust. Creating clear guidelines for data collection and usage should factor into the use of AI-driven systems as well as transparency. Likewise, clients should be educated on the risks associated with the data collection and storage and take the appropriate measures in securing this information.

Predictive Maintenance

The decision-making processes used by AI systems are not always clear. The question of ethical accountability arises when the AI-driven system makes an incorrect prediction. This could be in the form of unnecessary maintenance or not predicting an impending failure. Unnecessary maintenance wastes resources and creates downtime. Downtime may lead to losses in earnings or disrupt critical operations which could pose safety hazards. When an AI makes an incorrect decision, who is held responsible? The responsibility could fall on the engineer, the developer, or the client. The question of accountability should be addressed upfront with clients when AI-driven systems are employed by introducing guidelines for maintenance protocols related to AI decisions. Engineers must ensure that AI systems that use predictive maintenance are reliable and will not fail to detect critical issues. It is an engineer's responsibility to prevent potential harm that may occur from a malfunction or misinterpretation of AI-driven systems when handling critical infrastructures.

Another potential impact of the predictive maintenance feature is the impact on employment. Maintenance workers, plant engineers, and technicians may find their roles diminished when AI-driven systems are put into effect. While the end result for the building may be reduced costs, for the worker, it could mean job displacement. Arguably, with the invention of the car, there are less carriage drivers and farriers as well; and the economy adjusts accordingly. Still, where AI-systems are employed at scale, the effect on the employment, even if only temporary, is worth considering.

Finally, algorithms used for preventative maintenance should be trained and audited for any potential biases. If users enter flawed data inputs, the result could be a biased maintenance practice where certain areas of a building are prioritized over others. The key here is to determine if the bias is warranted or not. In a healthcare setting, critical areas operating 24/7 may naturally be more biased over administrative spaces which are far less critical. The level to which this is necessary depends on the context but protocols for auditing models for system bias should be standard.

Sustainability

In economics and other complex problems, there are no solutions, only trade-offs. The previous sections discussed system bias and energy sustainability. AI systems that make decisions on material use may be biased towards a variety of factors. AI systems that are capable of selecting materials for building systems could pose potential ethical concerns regarding the use of sustainable materials and reducing waste.

AI systems could prioritize a cheaper option which is ideal for the short term but is not sustainable long-term. That means factoring in the lifecycle of the material and the eventual disposal or recycling of the material. Likewise, reducing waste in one area of the facility may mean offsetting that waste in other areas of the building if the data is biased. A lack of transparency in the AI's decision-making process could result in a failure to uphold a facilities goal for sustainability.

One way to address this would be to create a hierarchy of sustainability goals which the AI could be trained in. These goals could be energy efficiency, waste reduction, reduce carbon footprint, etc. Training the model to provide clear explanations for its decisions would also help determine if the AI is prioritizing the facilities sustainability goals.

Retrofitting and Upgrades

Engineers employing AI driven systems which propose retrofits and upgrades to existing systems must consider the financial impacts of the decisions as well as the effects on building integrity.

The engineer must ethically weigh the costs and benefits when implementing AI-driven upgrades from a financial aspect. This is particularly the case for small businesses which may not have the financial resources to invest in more advanced systems. It might be the case that a potential system type could save more energy but if it cost the owner more money to the extent that it handicaps their bottom-line, the upgrade may not be worth the trouble. Furthermore, as might be the case with proprietary systems, it is important that the engineer recognizes any system bias towards more profitable markets or equipment that may not be the best choice for the client.

With all the pressure to modernize buildings, it can be easy to lose focus on concepts like preservation of existing structures. Historical buildings may be compromised from modern equipment suggested by AI. Upgrades and retrofits should be minimally invasive to be respectful of a building's historical or architectural significance.

One last consideration involves the associated environmental impact of upgrading equipment. New equipment must be manufactured, transported, and installed. All these activities use resources and energy. Furthermore, the disposal of old systems may contribute to waste if not managed properly.

If the engineer is concerned with environmental impact these consequences should be factored into the decision to move forward on an AI proposed upgrade or retrofit.

Human Oversight

As discussed in previous sections, AI-driven systems should ideally make decisions that align with safety standards, ethical principles, and safety standards. There will be a natural inclination to potentially over rely on these systems and trust them implicitly. This will especially be the case as these models become more sophisticated. The main crux with AI-driven systems is that technology can never replace human judgment. AI-systems will generally lack the moral and contextual understanding that a human possesses.

While technology can perform higher levels of computation and faster decision making, that is not necessarily indicative that the system management is ideal for safety-critical decisions. In such cases, human oversight should remain in place. Engineers along-side clients and developers must establish protocols for accountability if and when an AI-system fails. The goal is to ensure that there is no ambiguity about who is responsible for correcting issues when they do arise and to mitigate risks. Human oversight can take the form of regular audits or real time monitoring.

Conclusion

AI-driven systems already have expansive and proven effectiveness in optimizing building automation. While these systems have yet to be enacted at a large scale, at present, there is a growing demand for implementing AI. Engineers should understand the practical and ethical challenges associated with integrating AI within the building system. In order for AI-driven systems to operate ethically, they must take into account safety, transparency, sustainability, accountability, bias, privacy, and security concerns. Ultimately, AI should be thought of as a tool rather than a replacement for human oversight. By maintaining a strong human presence in the final decision making process, clients can reap the benefits of AI without sacrificing ethics or suffering hidden costs.

References

Bouabdallaoui Y, Lafhaj Z, Yim P, Ducoulombier L, Bennadji B. Predictive Maintenance in Building Facilities: A Machine Learning-Based Approach. Sensors (Basel). 2021 Feb 3;21(4):1044. doi: 10.3390/s21041044. PMID: 33546418; PMCID: PMC7913483.

Bowell, Andre. (2023, October). The Perfect Way to Smooth You Noisy Data. Medium. https://towardsdatascience.com/the-perfect-way-to-smooth-your-noisy-data-4f3fe6b44440

Dhamaodharan, Balaji. (2021, March). Optimizing Industrial Operations: A Data-Driven Approach to Predictive Maintenance Through Machine Learning. International Journal of Machine Learning for Sustainable Development

Gillingham, Kenneth, Rapson, David, & Wagner, Gernot. (2015, September). The Rebound Effect and Energy Efficiency Policy. Review of Environmental Economics and Policy

Goodfellow, Ian, Bengio, Yoshua, & Courville, Aaron. (2016). Deep Learning. MIT Press. (97-104, 116-118, 128)

Strubell, Emma, Ganesh, Ananya, & McCallum, Andrew. (2019, June). Energy and Policy Considerations for Deep Learning in NLP. https://doi.org/10.48550/arXiv.1906.02243