DOI: https://doi.org/10.48009/3_iis_2022_123

EMMA - The ethical multimodal, modeling and adaptive AI system for fairness in building tenant comfort

Christopher B. Davison, *Ball State University, cbdavison@bsu.edu*
David Hua, *Ball State University, dhua@bsu.edu*
Edward Lazaros, *Ball State University, ejlazaros@bsu.edu*
Allen Truell, *Ball State University, atruell@bsu.edu*
Brianna Bowles, *Ball State University, blbowles@bsu.edu*
Erin Boomershine, *Ball State University, emboomershin@bsu.edu*
Colin Allen, *University of Pittsburgh, colin.allen@pitt.edu*

Abstract

Artificial Intelligence (AI) is one of the most impactful and disruptive technologies of the current time. AI technology is developing in almost every aspect of human civilization from warfighting to toys. In this study a prototype AI (EMMA) in presented. The system is deployed in a Heating, Ventilation, and Air Conditioning (HVAC) infrastructure at Ball State University (BSU). EMMA is a research testbed for examining usability, fairness, and explainability in AI.

Keywords: AI, Machine Wisdom, Ethics

Introduction

AI continues to expand as it is deployed in almost every aspect of human living including financial markets, transportation/logistics, warfighting, and even benign spaces such as entertainment and video games. With this continued and explosive growth of AI, the following research questions arise:

1) Is the AI's decision explainable or understandable to those stakeholders impacted by the AI's decision-making? (XAI)
2) Can the AI provide meaningful and understandable explanations of the decision-making process and the fairness of the process? (FAI)

The purpose of this research article is to present the authors' AI prototype (EMMA), and describe the suitability of EMMA as a research infrastructure for exploring machine ethical reasoning, fairness (FAI), and explainability (XAI). Several theoretical constructs are presented including the EMMA Fairness Model.

The authors begin the paper with a review of the literature which discuses AI and the implementation of AI in HVAC systems. HVAC concepts such as comfort, comfort profiles, and comfort sensing are presented and explained. AI concepts such as fairness and explainability of AI are then presented. Following the literature review, the EMMA system is presented along with the authors' experimental designs within EMMA within the domain of AI fairness and explainability.

Literature Review

HVAC, BEMS

A building energy management system (BEMS) is a practical method used for monitoring and controlling energy consumption within buildings by reducing heating and cooling energy (Hannan et al., 2018). This is achieved through software by optimizing and integrating passive heating and cooling systems. Implementing a BEMS system that utilizes the Internet of Things (IoT) technology to manage multiple building facilities results in energy reduction, decreased labor costs, extended building lifetimes, and increased user comfort is termed a smart BEMS (Park et al., 2020).

In order to construct a smart BEMS, IoT devices are affixed to the building's surface, in high proximity to the user and the environment. The IoT sensors and control interfaces are then installed on the surface of the facility to collect user and environmental information and to intelligently control the facility based on this information (Park et al., 2020). Large amounts of data are then obtained by analyzing the user and environmental information from the IoT devices over a period of time, which in turn allows energy-saving methods to be applied to the building's HVAC system. An optimal HVAC environment for occupants, and reduction of energy are the primary goals of the smart BEMS. Therefore, the HVAC system references certain guidelines based on the data analysis. The most widely used technology for such data analysis today is the use of artificial intelligence (AI) (Park et al., 2020).

There are two different types of BEMS defined by Park et al. (2020): new BEMS and existing BEMS. New BEMS incorporate smart IoT technology into the construction of a new building whereas existing BEMS are implemented into a newly remodeled building. BEMS in new buildings are essentially low-cost during the pre-planning stage of construction. In addition to the low cost, other advantages, including cost savings, are realized through the selection and application of high-efficiency building equipment, assembling the latest IoT-based HVAC, and integrating the management system. However, implementing a BEMS in an existing building results in some increased costs due to the need for the replacement of expensive components and hiring additional building managers to apply the smart IoT system. The most reasonable option to decrease energy waste in such situations would be to install and apply light and low-cost IoT devices within the building to monitor environmental and energy usage information (Park et al., 2020).

Hannan et al. (2018) identified several crucial characteristics for BEMS systems such as (1) energy consumption management, (2) trending and benchmarking, (3) fault detection and diagnosis, (4) measurement and verification, (5) model-based BEMS control, and (6) construction operation building information exchange. Energy consumption management refers back to the two different types of BEMS (new and existing) because the primary goal of IOT-based BEMS is to save energy. Therefore, energy efficiency depends on renovations of existing buildings and the construction of new buildings. Renovating old buildings saves energy and materials compared to new construction, while also reducing emissions and waste, indicating that renovations better address climate change mitigation for environmental development and it is the preferred BEMS implementation method (Hannan et al., 2018).

The second characteristic of BEMS includes trends and benchmarks. Trends examine the buildings' energy bills and usage to define performance. This is particularly useful when identifying areas of extreme usage and making weather and maintenance adjustments as necessary. Benchmarking provides additional information regarding the building's potential efficiency in process, places, and outputs of energy usage.

Fault detection and diagnosis (FDD) automatically senses and segregates breakdowns in BEMS, aiming to safeguard a system from further damage/loss. FDD methods have been classified as model-based FDD, signal-based FDD, knowledge-based FDD, active FDD, and hybrid FDD. Although FDD has known limitations in data management, cost, and scalability; research has proven size, maintenance, and calibration of HVAC systems play a significant role in the reduction of energy waste (Hannan et al., 2018).

The fourth characteristic of BEMS includes the measurement and verification (M&V) of building data. This is utilized in order to appraise energy efficiency, energy demand, and consumption of various building equipment. Measurement is completed through the evaluation of a complete system or part of a system. A successful M&V plan should include operating hours, existing controls, light level, site selection, and HVAC effects (Hannan et al., 2018).

The fifth characteristic is a model-based BEMS control system, where the control parameters of the building are conveyed mathematically into a BEMS. Existing control strategies (on-off control, PID control, and rule-based control) have presented limitations in adjustment and therefore, advanced control strategies (intelligent control and advanced fuzzy logic control) have been introduced. To compensate for these limitations, the model-predictive control (MPC) approach consists of problem formulation, a control architecture, and various implementation types.

The final characteristic includes the construction operation building information exchange (COBie). COBie is defined as the particular set of building information that can be delivered to owners/operators in a standard manner. Examples include developing a building information transformation model, building information management, and facilities information management (Hannan et al., 2018).
Research conducted by Yan et al. (2020) identified AI applications in buildings include the following information: (1) indoor environmental detection and control - temperature, humidity, and air quality, (2) efficiency of building multi-energy utilization, and (3) predictive accuracy of forecasting in buildings - load forecasting, subsystem (e.g., HVAC, lighting) performance forecasting and building structure safety. In this work, the researchers identified four stages, or levels, of which the AI in the buildings could be classified: *basic bundled*: where similar smart technologies are grouped based on common interfaces, *automated:* allowing the applications to have higher automaticity and the systems and facilities are capable of interconnection and anticipation of needs, *intuitive:* where applications learn independently and acclimate services to context, and *sentient:* meaning the applications are able to recognize and meet the needs of the building through predictive analytics. Each level represents a significant increase in AI complexity and integration into the HVAC system.

Comfort and Sensing of Comfort

Cities, and the world as a whole, will be layered with sensing and actuation, according to futurists, resulting in a Smart World (Stankovic, 2014). Buildings will be designed employing modern, integrated materials, sensors, electronics, and networks that are interfaced with computerized systems as part of this concept (Bowerman et al., 2000). Sensors, electronics, and nano-technological materials have already been installed in many buildings to make them more eco-friendly and energy-efficient (Bradshaw, 2006; Zeiler et al. 2011). In the recent decade, home appliance automation has been proposed to increase efficiency and improve user experiences (Lu & Whitehouse, 2010) and frequently integrates the use of smartphones as sensors (Trappl, 2015). These advancements in sensor system development and implementation will generate a huge qualitative shift in how we work, live, and do things (Stankovic, 2014).

A building that monitors and learns the conditions of vital infrastructures, such as energy usage, water, power, heating, cooling, and communication systems, can better maximize its resources, prepare preventive actions, and monitor security aspects while improving occupant comfort (Bowerman et al., 2000). Thermal comfort is a critical aspect of the indoor environmental quality of a building (Frontczak & Wargocki, 2011). However, considerable differences in physiological and psychological comfort among people might make finding an ideal temperature for everyone in a particular location problematic. Furthermore, it can be difficult to identify crucial parameters in real-time (Olesen, 2004). Rather than taking individual preferences into consideration, most BEMS simplify the work by employing standard suggestions such as ASHRAE Standard 55. (Jazizadeh et al., 2014).

AI, FAI and XAI in BEMS

While there is no universally agreed definition of fairness, the authors of a recent survey article proposes the following general definition: "the absence of any prejudice or favoritism toward an individual or a group based on their inherent or acquired characteristics" (Mehrabi et al., 2021, p. 11). This definition provides a basic outline for fairness but does not directly address how this definition might change depending on the situation and context. In order to remedy this, multiple definitions of fairness that are directly tied to machine learning and fairness have been created and presented as a "taxonomy of fairness" ((Mehrabi et al., 2021, p. 1). Those definitions include Demographic Parity, Conditional Statistical Parity, Equalized Odds, Equal Opportunity,Treatment Quality, Test Fairness, Subgroup Fairness, Fairness Through Unawarness, Fairness Through Awareness, Counterfactual Fairness, and Fairness in Relational Domains. Related to the goals of the EMMA system and HVAC fairness, the following definitions of fairness from Mehrabi et al. (2021) appear to be applicable, observable (ie., tracked through EMMA's data collection mechanisms), and measurable: fairness through awareness, fairness in relational domains, and the categorical classification of subgroup fairness.

Algorithms using fairness through awareness can be perceived as fair if the algorithm gives similar predictions to similar individuals (Verma & Rubin 2018). This definition directly relates to the working hypothesis of this project, that fair outcomes are enhanced when EMMA presents explanations about other users' needs and preferences. As certain attributes such as age, gender, ethnicity (all Fair Housing Administration and Equal Credit Opportunity Act protected attributes) and medical conditions (HIPAA protected information), are collected and stored in the EMMA database systems through voluntary self-disclosure, decision-making bias can be analyzed in a post-hoc fashion.

Fairness in relational domains is the idea that in order for an algorithm to be fair it must capture relational structure through "not only taking attributes of individuals into consideration but by taking into account the social, organizational and other connections between individuals" (Mehrabi et al., 2021, p. 12). This definition implies that algorithms can not only capture relationships present within data but also then use these relationships in order to compare other attributes to produce a similar result. In the context of EMMA, fairness in relational domains allows the different roles of tenants (e.g., student, faculty, custodian, security) in buildings to be factored into the fairness decision. Without taking these tenant roles into account when discussing and deciding fairness, there is a potential for domain bias within the EMMA system.

To better understand the definitions above researchers have created categorical classifications, one of those classifications being subgroup fairness (Mehrabi et al., 2021). Subgroup fairness is described as intending to "obtain the best properties of the group and individual notions of fairness … It picks a group fairness constraint … and asks whether this constraint holds over a large collection of subgroups" (Mehrabi et al., 2021, p. 13). This subgroup definition contains the basic definition of fairness, fairness in relational domains and fairness through awareness, and outlines how the data is compared in order to produce a fair result. In the context of EMMA these subgroups could range from transient students to stationary students, or from students to faculty. Subgroup fairness allows researchers to understand the weight, if any, that each of these groups carry when making a decision. Other subgroup fairness classifications could be daytime or nighttime tenants or subgroupings by virtue of length of stay or office location.

Researchers have developed methods that aim to directly satisfy each proposed definition of fairness. One of these methods is fairness regression which consists of assessing three fairness penalties. These penalties are referenced as the *Price of Fairness* (POF) and are used to account for what is described as accuracy fairness tradeoffs (Mehrabi et al., 2021). Each POF penalty is based upon the classification and categorical placement of fairness being assessed. Those POF categories are individual fairness, group fairness as well as hybrid fairness, which are defined below. Researchers have defined the individual price of fairness as "for every cross pair (x,y) ∈ S1, (x′ ,y′) ∈ S2, a model w is penalized for how differently it treats x and x′ (weighted by a function of |y − y′|) where S1 and S2 are different groups from the sampled population" (Mehrabi et al., 2021, p. 16). This POF penalizes algorithms that treat the initial value or input differently than the negation of that same value, ensuring fairness for each value entered. The group price of fairness is defined by researchers as "On average, the two groups' instances should have similar labels (weighted by the nearness of the labels of the instances)" (Mehrabi et al., 2021, p. 17). Group fairness penalizes algorithms that do not compare groups equally based on a set of both instances and labels.

Fairness in relational domains and fairness through awareness both focus on taking individual attributes into account in order to produce a decision. These individual attributes range from user attributes, relational domain attributes and subgroup attributes. In order for EMMA to make a fair decision regarding its users, all individual attributes must be considered and applied correctly/consistently. For instance, when dealing with one user, EMMA is able to capture all associated attributes: demographics, temperature preference and room location. EMMA can then use these attributes to compare other users' requests. If two users are in the same room, they may be considered in the same group, but not the same subgroup as one individual may have a medical condition while the other does not. The concept of fairness is exemplified by EMMA treating the two individuals fairly but accommodates on behalf of medical conditions. The user without a medical condition present will still be treated fairly as EMMA provides them with an explanation as to why their request could not be effectuated.

In the setting of a BEMS, questions of fairness, bias, and inclusion occur in ways that are not always obvious. For example, due to a mix of clothing codes and fashion, males are frequently more densely clad than females in the same spaces, resulting in distinct temperature and humidity preference distributions. It is also possible that different types of building tenants (e.g., custodians, students, faculty, and administrators) spend different amounts of time in the building, have various activity levels, and leave and re-enter at different frequencies, resulting in differing perspectives on the best building climate settings. Distinct socioeconomic distributions result in variable expectations and preferences, as well as diverse probabilities of complaining about discomfort in these various roles. How can and should these opposing desires be assessed against other criteria such as the building's overall efficiency and the heating and cooling's environmental impact? The authors of this article suggest collecting data on how different EMMA users engage with the building and express their preferences in order to assess correlations between actions (e.g., time spent in the building) and preferences, and then developing algorithms that fairly balance the tradeoffs considered.

Machine Ethics, Machine Wisdom, and Machine Phronesis

Ethics is a field of humanistic theory concerned with determining moral standards and resolving moral dilemmas. Metaethics is the study of the fundamental nature of moral ideals and how they might be justified philosophically. Ethics and morality were once thought to be solely the domain of human agents. Machines' increasing complexity and operational autonomy have led to a growing recognition that they will require some ethical capabilities (Picard, 1997; Allen, Varner & Zinser 2000; Georges 2003). As a result, over the last two decades, the area of Machine Ethics has emerged (Arkin 2009; Wallach & Allen 2009; Anderson & Anderson, 2011; Lin, Abney, & Bekey 2011). Machine ethicists, on the other hand, aren't interested in creating artificial meta ethicists; rather, they're interested in figuring out how to create machines that can make decisions based on ethical principles set by their creators (see Bringsjord & Taylor, 2011, for more on "divine command roboethics"; also Georges 2003). While some academics contend that Machine Ethics' whole premise is flawed (Yampolskiy, 2013), others have developed broad frameworks for artificial moral beings (Anderson, Anderson & Armen, 2006; Arkin, 2009; Wallach, Franklin & Allen 2010). The EMMA researchers will contribute to the growing trend of building working prototypes (Dennis et al. 2014; Trappl 2015).

The EMMA researchers are not attempting to tackle the difficult task of constructing human-level AMAs with emotional and cognitive abilities (Wallach & Allen, 2009; Sparrow, 2009). An *ethical governor* (Arkin, 2009; Kinne & Stojanov, 2016)—a distinct module to perform ethical judgment and limit the possibilities that an autonomous system can choose—is a more tractable solution within existing technology that is embraced by the authors. Machine ethicists disagree about which metaethical frameworks can and should be used in AMAs (see Tonkens, 2009; Arkin, 2010; White, 2015; Arkin, 2010; Leben 2018). The researchers on this project use a rule-based utilitarian ethical framework for EMMA. Minorities and other disadvantaged groups have been claimed to benefit from rule-based utilitarianism. The rule-based approach will allow scientists to reflect essential BEMS principles such as tradeoffs between individual needs and preferences, consideration for those most affected by fairness and power imbalances, total societal goods, and machine self-preservation. The result is a model that represents and parameterizes these and other parameters based on data from a variety of physical, environmental, and human inputs. The principles serve as the foundation for EMMA's FAI decisions to be explained (see the EMMA Fairness Model in Figure 3).

These variables will help explain how EMMA responds to user requests or complaints of discomfort. For example, if changes will stress the HVAC system (BEMS self-preservation), have an unacceptable environmental impact (global values), or have a predicted detrimental impact on specific classes of building tenants, a user report of being too hot might not result in any change (fairness). The EMMA researchers may then evaluate whether elements are more convincing to different tenant categories if the decision can be communicated to the user in terms of these deconstructable aspects.

The goal of the EMMA system is for wisdom to be embedded in the entire socio-technical system that human actors and AI inhabit. This amounts to engineering a sort of practical wisdom into the system. The researchers term this *machine phronesis* (borrowing from, but not necessarily endorsing, Aristotle's thoughts on practical wisdom) or *wisdom by design*. Utilizing a bounded construct (e.g., a building HVAC system) provides a known parameterized testbed in which to experiment.

EMMA System

Building Energy Management Systems (BEMS) or Smart Buildings are used to monitor and control the electrical and mechanical systems in buildings, including heating, ventilation, air conditioning (HVAC), and lighting, with the objective of providing the expected level of comfort in addition to optimizing energy consumption. These systems are traditionally operated based on feedback from sensors (including temperature, CO2, and relative humidity sensors) in addition to time schedules in some buildings. The current focus on integrating AI into BEMS tends to focus on cost savings and, to a lesser degree, predictive resource consumption and occupant comfort (Wright et al. 2002). However, no ethical, fairness, or inclusivity considerations are factored into BEMS, with the exception of energy-saving resource management decisions (a societal good).

Purpose of the EMMA Infrastructure

The objective of this research infrastructure is to implement ethical reasoning into a BEMS AI and study whether it will enhance the fairness and explainability of the BEMS's decision-making through the incorporation of models of tenant behaviors, tennant input, adaptive algorithms, sensorization, and AI-provided explanations and feedback. In its current state, EMMA is the interface with the BEMS (currently provided by Automated Logic Corporation), with BEMS sensors and data streams providing data to EMMA. Tenants provide EMMA with comfort profile preferences through the EMMA mobile application. The target building for the EMMA prototype is the Applied Technology (AT) building (3 floors and a total area of 7,821m^2) at Ball State University in Muncie, Indiana. Currently, EMMA is operational and the researchers are conducting experiments.

The researchers will investigate the use of behavioral data and user-supplied comfort profiles and ratings to enhance AI ethical factors including fairness, transparency, explainability, and inclusivity within the Smart Building context. The user data will support explainability on two levels: machine-to-human explainability and machine-to-machine explainability.

Human-machine explainability includes human-centric approaches that account for context, user experience, cultural background, preferences, and other factors. In order for the explanation to be useful, the AI must explain its decision in a fashion that is understandable by the user (Creel, 2021).

Machine-machine explainability recognizes that machines (actual hardware or software agents, often acting as surrogates or on behalf of humans) may also require an explanation for AI outcomes in highly-connected and instrumented Internet of Things (IoT) spaces. These two explainability models are quite different in most respects (e.g., language, speed, delivery modality), but they share the core principle that successful explanation entails providing simplified descriptions of complex phenomena (Mittelstadt et al., 2019).

Design and Architecture

One of the project's main goals is to create a system (EMMA) that can model, anticipate, and intervene in energy usage and behavior while maintaining justice, ethics, and explainability. The PIs plan to employ open-source software whenever possible for simplicity of use and integration into a genuine system, as well as to keep expenses down. Currently, the system communicates with Automated Logic Corporation's (ALC) BEMS.

Figure 1 depicts the essential components of the EMMA System back-end system. Figure 2 depicts the current EMMA front-end system, which displays how users of the EMMA App provide feedback to the system in two dimensions (temperature and humidity), rather than asking for a specific temperature. The App is also meant to provide feedback to users, as demonstrated; however, the user prompts, surveys and other features proposed in this research article are not visible.

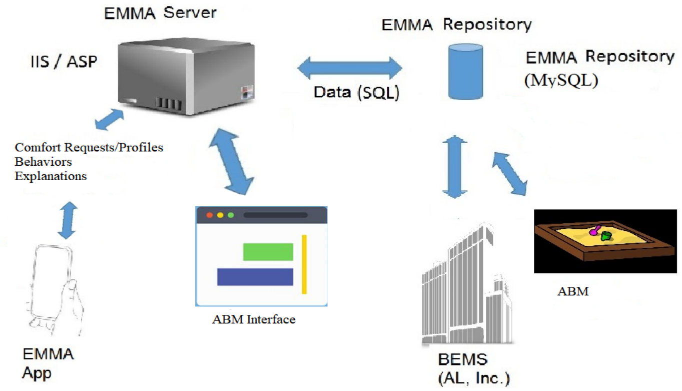
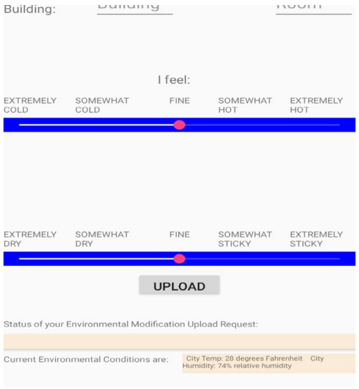


Figure 1. Prototype EMMA Back-End System



Status of your Environmental Modification Upload Request:

Current Environmental Conditions are: City Temp: 28 degrees Fahrenheit City Humidity: 74% relative humidity

Figure 2. Prototype App

EMMA's Fairness (FAI) Model

EMMA's fairness decision-making algorithms consider machine imperatives including survival and maintenance. If a user's request puts too much demand on the HVAC system (for example, requests that cause frequent HVAC on-off activity), the decision-making algorithms account for this, and the user may receive a "no" or "partial fulfillment" response to their comfort request. Current BEMS discreetly fulfill or refuse users' requests, making it difficult for consumers to understand or even be aware of the response or its fairness. The following is how EMMA will inform users about FAI's impact.

The capability to actively communicate itself to users will be unique to EMMA's ethical decision-making. Whether a user receives a "no," "yes," or "partial fulfillment" response to their request, EMMA will ask whether the user wants more information about the choice. To arrive at a fair decision, the EMMA Fairness Model (provisionally sketched in Figure 1) employs an algorithm to balance decision-making parameters and system attributes (potentially including but not limited to those depicted in the quadrants of Figure 1). A time-sensitive smoothing algorithm will deliver real-time responses to user requests given resources available, taking into account parameters that will be determined throughout the project.

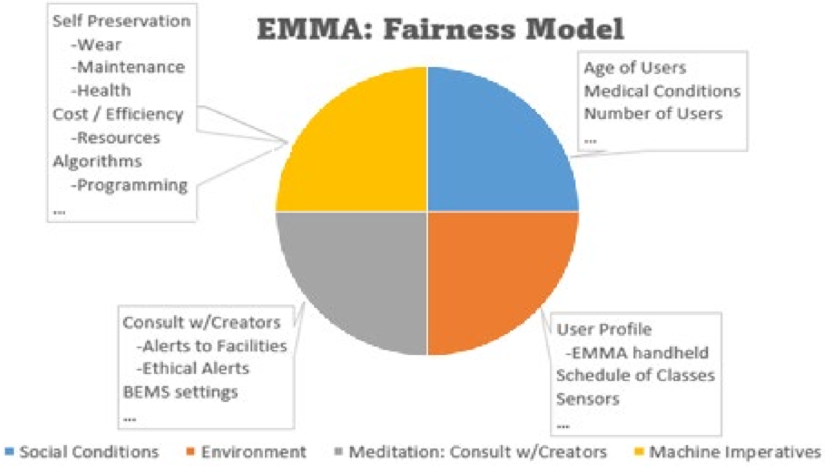


Figure 3. The EMMA Fairness Model (Conceptual Model).

If the user wants more information, EMMA will respond with more information. This is how the XAI notion of "generations of explanations" is put into practice (Xu et al., 2019, p. 9). On successive cycles, each level of explanation becomes more thorough and hence more technical. EMMA moves to the context of machine-to-machine explanation when the user is dissatisfied or the details are beyond their comprehension, where the EMMA App, working independently from the backend server, is a machine surrogate for the user's best interests that can perform the additional context needed to conduct a fairness check and report potential bias to the user.

Why AI and HVAC?

Building Energy Management Systems (BEMS) or *Smart Buildings* regulate and control electrical and mechanical systems in buildings, such as heating, ventilation, air conditioning (HVAC), and lighting, with the goal of ensuring desired levels of comfort while reducing energy usage. Sensory feedback (including temperature, CO2, and relative humidity sensors), as well as schedules in some buildings, have typically been used to operate these systems. The present focus on incorporating AI into BEMS is on cost reductions, with a lower emphasis on predicting resource usage and occupant satisfaction (Wright et al. 2002). BEMS, on the other hand, does not take ethical, justice, or inclusivity into account, apart from energy-saving resource management decisions (a societal good).

The EMMA researchers are exploring the use of behavioral data and user-supplied comfort profiles and ratings to strengthen AI principles such as fairness, transparency, explainability, and inclusivity within the Smart Building context in this joint proposal from Ball State University (BSU) and the University of Pittsburgh (Pitt). Explainability will be supported on two levels using user data: machine-to-human explainability and machine-to-machine explainability. Human-machine explainability encompasses human-centered techniques that consider background, user experience, cultural background, preferences, and other ethical considerations. Machine-machine explainability recognizes that in highly linked and instrumented IoT (Internet of Things) settings, machines (real hardware or software agents, frequently acting as a proxy or on behalf of humans) may also demand an explanation for AI outcomes. In most ways (e.g., language, speed, and delivery medium), these two explainability models are significantly different, yet they both share the essential idea that successful explanation includes offering simpler explanations of complicated occurrences (Mittelstadt et al. 2019).

The EMMA system and applications have real-time access to BSU's BEMS. Tenant behavior, tenant fairness considerations (e.g., medical conditions, disabilities, as well as other equality and fairness factors), tenant demographics (e.g., age, gender identity), social conditions, resource conservation (in terms of monetary and societal cost), and the HVAC system's own imperatives are all sources of EMMA data (survival, well-being, mechanical condition). This information is used to create a complete framework for BEMS operation and control. This project builds on an existing operational prototype for ethical AI on the BSU campus, which will be enhanced and evaluated for FAI utilizing XAI approaches (e.g., sensitivity analysis, visualization, rationalization).

Future Research in the EMMA Infrastructure

The researchers on this project envision an evolution of EMMA as an AI-surrogate platform. The concept is the handheld EMMA will act as a champion for the user in terms of XAI and FAI. This is an example of fighting fire with fire (in this case fighting AI with AI) or at least leveling the playing field with equivalent technologies.

While the user may not remember the frequency of dates of denials or partial request fulfillment, the surrogate will always remember (network or handheld/wearable failure notwithstanding).

The EMMA handheld surrogate can intercede on behalf of the user and demand accountability (XAI) from the AI with which the user is interacting.

As an example in the context of the instrumented space at BSU: a user enters a building once a month for a meeting. This user's comfort request is denied more often than not. The user may not notice, may not care, or just forget. However, EMMA (the handheld surrogate) can demand to know the percentage of denied requests, frequency of denied requests, and demographics of those that are denied. From that and other requested data, EMMA can determine if bias exists in the decision-making AI.

Funding and IRB Protocol

The Machine Wisdom Project (of which EMMA is a subproject) is funded by a grant from the Templeton World Charity Foundation Diverse Intelligences program to Colin Allen, Department of History & Philosophy of Science, University of Pittsburgh, and a sub-award to Chris Davison, Center for Information and Communication Sciences, Ball State University, Indiana. For more information regarding The Machine Wisdom Project, see: https://themachinewisdomproject.org/

The research is covered under an in-process BSU IRB protocol 1884089-1: Machine Wisdom: Wise Artificial Intelligence in Smart Buildings.

Conclusion

In this research article, an emerging, disruptive technology: EMMA, The Ethical Multimodal, Modeling and Adaptive AI System was presented. EMMA is deployed with the bounded context of a campus HVAC system. This provides a research testbed to explore fairness in AI and explainability in AI.

The EMMA Fairness Model consists of social conditions, environmental conditions, machine imperatives and operational parameters. The EMMA system examines these factors which may consist of comfort requests, user demographics, machine health, and others, to arrive at a decision on temperature and humidity in a building area. The EMMA system will explain to a user how the decision was reached.

The HVAC system appears to be an ideal socio-technical testbed for practical wisdom (i.e., machine phronesis) in AI. The system is sensor and actuator rich, has many users of varying types (e.g., students, custodians, staff, faculty) with different comfort preferences, and is bounded and parameterized. This provides a unique testbed environment in a little-explored area of AI research: HVAC.

References

Allen, C., Varner, G., & Zinser, J. (2000). Prolegomena to any future Artificial Moral Agent. *Journal of Experimental and Theoretical Artificial Intelligence 12*, 251-261.

Anderson, M. & Anderson, S.L. (Eds.) (2011). *Machine Ethics*. Cambridge University Press.

Anderson, M., Anderson, S. L., & Armen, C. (2006). An approach to computing ethics. *IEEE Intelligent Systems*, 56-63.

Arkin, R. C. (2009). "Governing Lethal Behavior in Autonomous Robots." Boca Raton Fla.: CRC Press.

Arkin, R. C. (2010). The case for ethical autonomy in unmanned systems. *Journal of Military Ethics*, *9*(4), 332-341.

Bowerman, B., Braverman, J., Taylor, J., Todosow, H., & Von Wimmersperg, U. (2000, September). The vision of a smart city. In *2nd International Life Extension Technology Workshop, Paris* (Vol. 28).

Bradshaw, V. (2006). The Building Environment: Active and Passive Control Systems. River Street, NJ: Wiley.

Bringsjord, S. & Taylor, J. (2011). *Introducing divine command roboethics*. In Lin, P., Abney, K.,Bekey, G. (Eds.) *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge: MIT Press; pp. 85-108.

Creel, K. A. (2021). Function and User-Satisfaction in Explainable AI. Overcoming Opacity in Machine *Learning*, 3.

Dennis, L., Fisher, M., Slavkovic, M., Webster, M. (2014). *Towards Autonomous Robotic Systems: 14th Annual Conference.* TAROS 2013, Oxford, UK, August 28--30, 2013. In Ashutosh Natraj and Stephen Cameron (Eds.), Springer, 433-445

Frontczak, M., & Wargocki, P. (2011). Literature survey on how different factors influence human comfort in indoor environments, *Building and Environment*, *46*, 922-937.

Georges, T.M. (2003). *Digital soul: Intelligent machines and human values*. Cambridge: Westview Press.

Hannan, M. A., Faisal, M., Ker, P. J., Mun, L. H., Parvin, K., Mahlia, T. M. I., & Blaabjerg, F. (2018). A Review of Internet of Energy Based Building Energy Management Systems: Issues and Recommendations. IEEE Access, 6, 38997–39014. https://doi.org/10.1109/access.2018.2852811

Jazizadeh, F., Ghahramani, A., Becerik-Gerber, B., Kichkaylo, T., & Orosz, M. (2014). User-led decentralized thermal comfort driven HVAC operations for improved efficiency in office buildings. *Energy and Buildings, 70*, 398-410.

Kinne, E., & Stojanov, G. (2016, March). Grounding Drones' Ethical Use Reasoning. In *2016 AAAI Spring Symposium Series.*

Leben, D. (2018). Ethics for robots: *How to design a moral algorithm*. Routledge.

Lin, P. Abney, K. & Bekey, G. (Eds). (2011). *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge: MIT Press.

Lu, J., Birru, D., & Whitehouse, K. (2010, November). Using simple light sensors to achieve smart daylight harvesting. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building* (pp.73-78). ACM.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1-35. https://doi.org/10.1145/3457607

Mittelstadt, B., Russell, C., & Wachter, S. (2019, January). Explaining explanations in AI. I*n Proceedings of the conference on fairness, accountability, and transparency* (pp. 279-288).

Olesen, B. W. (2004). International standards for the indoor environment. *Indoor Air*, *14*, 18-26.Peschiera, G., & Taylor, J. E. (2012). The impact of peer network position on electricity consumption in building occupant networks utilizing energy feedback systems. *Energy and Buildings, 49*, 584-590.

Park, S., Park, S., Choi, M.-in, Lee, S., Lee, T., Kim, S., Cho, K., & Park, S. (2020). Reinforcement learning-based BEMS architecture for energy usage optimization. *Sensors*, *20*(17), 4918. https://doi.org/10.3390/s20174918

Sparrow, R. (2009). Building a better WarBot: Ethical issues in the design of unmanned systems for military applications. *Science and Engineering Ethics*, *15*(2), 169-187

Stankovic, J. (2014). Research directions for the internet of things. *Internet of Things Journal, IEEE*, *1*(1), 3-9.

Tonkens, Ryan (2009). A challenge for machine ethics. Minds and Machines, 19 (3):421-438.

Trappl, R. (Ed.) (2015). *A Construction Manual for Robots' Ethical Systems*. Cham, Switzerland: Springer.

Verma, S., & Rubin, J. (2018). Fairness definitions explained. Proceedings of the International Workshop on Software Fairness - FairWare '18. https://doi.org/10.1145/3194770.3194776

Wallach, W. & Allen, C. (2009). *Moral Machines: Teaching Robots Right from Wrong.* New York: Oxford University Press.

Wallach, W., Franklin, S., Allen, C. (2010). A conceptual and computational model of moral decision making in human and artificial agents. *Topics in Cognitive Science 2*(3): 454-485.

White, J. (Ed.). (2015). *Rethinking Machine Ethics in the Age of Ubiquitous Technology*. IGI Global.

Wright, J. A., Loosemore, H. A., & Farmani, R. (2002). Optimization of building thermal design and control by multi-criterion genetic algorithm. *Energy and buildings, 34*(9), 959-972.

Yampolskiy, R. V. (2013). Artificial intelligence safety engineering: Why machine ethics is a wrong approach. In Philosophy and Theory of Artificial Intelligence (pp. 389-396). Springer Berlin Heidelberg.

Yan, B., Hao, F., & Meng, X. (2020). When Artificial Intelligence Meets Building Energy Efficiency, a review focusing on Zero energy building. Artificial Intelligence Review, 54(3), 2193–2220. https://doi.org/10.1007/s10462-020-09902-w

Zeiler, W., Houten, R., Boxem, G., Vissers, D., & Maaijen, R. (2011). Indoor air quality and thermal comfort strategies: the human-in-the-loop approach. In *Proceedings of the International Conference for Enhanced Building Operations.*