Tackling Algorithmic Transparency in Communal Energy Accounting through Participatory Design

Florian Cech
florian.cech@tuwien.ac.at
Centre for Informatics and Society, TU Wien
Vienna, Austria

ABSTRACT

Algorithmic transparency presents a significant challenge to system developers and users of algorithmic systems alike. Framing the problem as a 'wicked' one, this study tackles the issue of transparency in the EnerCoach energy accounting tool through presenting a situated ethnography of the algorithmic system and exploring the issues and challenges of model transparency and post-hoc explainability therein. By engaging stakeholders through participatory design methodologies, both a conceptual understanding of the problem and material solutions thereof are developed and evaluated. The findings show the promising potential of participatory design methodologies to elevate users to a 'critical audience', and the solutions co-created by the study participants for the challenge of algorithmic transparency. The results also highlight the complexity of the problem: transparency of algorithmic systems must be understood as a multi-facetted and highly contextual, 'wicked' problem that requires diverse methodological interventions to reach 'satisficing' solutions.

CCS CONCEPTS

• **Human-centered computing → Participatory design**; *Empirical studies in interaction design*; *Ethnographic studies*; • **Social and professional topics** → *Sustainability*.

KEYWORDS

Algorithmic Transparency, Algorithmic Literacy, Participatory Design, Critical Algorithm Studies

ACM Reference Format:
Florian Cech. 2021. Tackling Algorithmic Transparency in Communal Energy Accounting through Participatory Design. In *C&T '21: Proceedings of the 10th International Conference on Communities & Technologies - Wicked Problems in the Age of Tech (C&T '21), June 20–25, 2021, Seattle, WA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3461564.3461577

1 INTRODUCTION

Algorithmic transparency is rapidly becoming a crucial focal point of academic discourse in the fields of HCI, Science and Technology Studies (STS), and beyond. With the increasing digitalization of our current (information) society, algorithmic systems – including, but not limited to automated decision making, decision support systems and AI-based applications – can provide crucial support for complex tasks, but are often plagued by a lack of transparency and, subsequently, often leave those affected by the system in the dark about the processes and methods that lead to the final result. In some cases, this may even be an intended effect, leaving these systems as Black Boxes [27] by design that obfuscate the inner mechanisms to protect intellectual property and business interests of the owners. However, even for those aiming to make the inner workings of such systems transparent, the problem presents itself as multi-faceted and highly contextual. Neither is there a clear answer as to what constitutes adequate transparency for algorithmic systems in general nor for specific systems with different groups of affected users and stakeholders. Even transparency itself may take on different meanings on the spectrum between *an understanding of the internal processes, design decisions and general methodologies* to *post-hoc explanations of system outputs and decisions* [26]. Creating systems that provide exhaustive insights into their inner workings is often not possible or feasible, and may not even necessarily serve one of the main goals of transparency: to create systems that fulfill higher accountability standards [3, 38]. Finally, different stakeholders may have very different requirements towards the transparency of such systems, leading to competing incentives for different solutions. All these attributes qualify algorithmic transparency as what Rittel et al. call a *wicked problem* [28] with no *"definitive formulation"* or *"ultimate [...] test of a solution"* - a class of problems where every instance requires unique and iterative approaches to reach a *"satisficing"* solution [9, p.4].

To tackle this wicked problem, I investigate issues of algorithmic transparency within the growing field of civic technologies supporting sustainability, such as eco-feedback tools [11, 35] and energy accounting technologies [15]. Following Seaver's notion of *algorithms as culture* [31], I develop a situated ethnography of the collaborative energy accounting system *EnerCoach*[1]. Through this approach, I provide insights into the requirements of different stakeholders for transparency by framing the user activities related to transparency as sense-making activities [19, 22]. Based on these insights, the study produces material and conceptual solutions aiming to satisfy these requirements through participatory design methodologies [29] with the users and administrators of the system. By evaluating these interventions, I then derive learnings for the development of socio-technical measures to improve the transparency of algorithmic systems through stakeholder involvement.

[1] see https://www.local-energy.swiss/arbeitsbereich/energiestadt-pro/werkzeuge-und-instrumente/energiebuchhaltung.html

C&T '21, June 20–25, 2021, Seattle, WA, USA

ACM ISBN 978-1-4503-9056-9/21/06.
https://doi.org/10.1145/3461564.3461577

2 EXPLORATORY VIGNETTE AND PRIOR RESEARCH

As a part-time software engineer working for the company developing the EnerCoach energy accounting tool that this article focuses on, I was frequently confronted with the overwhelming opacity of the system. I present this exploratory composite vignette [34] to illustrate the challenges of improving the transparency of algorithmic systems from the vantage point of not only a researcher in critical algorithm studies, but also a software developer negotiating the trade-offs between the needs of users and stakeholders on the one hand, and technical requirements for developing such a system on the other. A significant part of my responsibilities as a software developer include troubleshooting implausible report results the system sometimes comes up with, which lead to support requests from the administrators of the system to determine whether or not the cause of this result is either a software error or a correct calculation that was based on incorrect data entered by the users. Many such cases end up being traced back to specific implementation details that sometimes originate within a single line of code. To resolve these issues, at least two steps of translating different levels of explanations are required: first, a generalized verbal description of still comparably technical reasons for the calculation result are to be communicated to the energy experts staffing the hotline, who have a good understanding of the energy accounting domain, but almost no knowledge of the technical side of programming such a system. Second, the hotline staffers must explain to the users not only *how* the system works and *what they can do about it* (e.g., enter data into the system differently), but also *why* the system was designed this way, and do so in a form that is reasonably understandable for the users. Considering this process, the potential for miscommunication and conflicts is obvious - and yet, as a researcher on algorithmic transparency I am aware that none of the standard measures of software development (such as better documentation or automated error-checking) to improve this process or prevent such instances from occurring seem to satisfy more than single, specific instances of such requests: no automated reasoning can explain every single decision made when specifying and implementing the system, particularly not given the diverse levels of both technical and domain specific literacy of its end users. Consequentially, these issues of troubleshooting a complex system like the EnerCoach tool exemplify the enormous challenge posed to users, administrators and developers by a lack of transparency, even if all of them are willing and able to cooperate on resolving them.

The specific example presented in this vignette fits within the larger corpus of research done in the nascent field of *critical algorithm studies* [16], concerned with a critical, interdisciplinary analysis of algorithmic systems. To this end, a broader interpretation of the terms 'algorithm' and 'algorithmic system' than usually employed within computer science alone is necessary. Without entering the fray of competing definitions across disciplinary boundaries in what Seaver calls *"terminological anxiety"* [31, pp.1], one solution is to follow his recommendation and treat algorithmic systems *as culture*, instead of as artefacts separate from, but embedded *in culture*. This approach hinges on the idea that algorithms are the *"manifold consequences of a variety of human practices"* [31, pp.4], co-created through the combined efforts of a variety of actors. For the example of EnerCoach, these practices include not just the stakeholders defining the system's parameters and the developer's actions of programming, deploying and maintaining the system, but also the users interacting with the system in a variety of ways, and shaping its nature and ongoing development through their feedback, questions and critiques. Consequently, the EnerCoach system is to be considered as an algorithmic system integrating both the social practices of the community of stakeholders involved in the system and the technical, material manifestations of those practices: a *"heterogeneous and diffuse socio-technical system"* [31, p.1] that requires both technical insight and ethnographic fieldwork to be studied in its entirety [30].

Transparency in algorithmic systems has been a focal point of attention of scholars in science and technology studies, computer science, and ethics in recent years. Numerous examples highlight the need for increased transparency or the harmful consequences of a lack thereof (e.g. [5, 8, 14, 18, 20]). Burell identifies three types of opacity in machine learning algorithms as *intentional* (for reasons of secrecy), as *technical illiteracy* and as an *inherent attribute* of the way machine learning algorithms operate [6]. Adding another layer, Geiger's landmark study on the use of algorithmic support technologies such as a bots in Wikipedia's organizational culture presents a fourth dimension: *"[...] the opacities in learning a particular institutional or organizational culture that is supported by algorithmic systems."* [14, p.4]. Beyond simply identifying different types of transparency, scholars also venture to elucidate the difficult relationship between transparency and accountability. Given the growing number of big data applications in governance and automated decision making in bureaucracies (e.g. [2]), holding these systems to account for their outputs and subsequent impacts on human lives require more than just transparency, as Kemper et al. point out: *"[...]without a critical audience, algorithms cannot be held accountable."* [18, p. 1]. Algorithmic transparency research must not, as they argue, leave the abstract *value* of transparency and the shape it may take for different target groups uncoupled from the practical implementation of concrete measures; instead, it must involve the potential audience and empower them to become a *critical* one in their engagement with developing these concrete measures. The fact that this plead for inclusion resonates so strongly with the roots of participatory design methodologies [29] also informed the methodological decisions made in this study, as explicated in the following section.

While transparency in algorithmic systems in general has garnered significant attention, looking towards transparency in energy accounting reveals a stark lack of research. After surveying the ample literature on energy and sustainability calculations, it becomes clear that the majority of research either focuses on tools and methodologies aimed at domain experts (e.g. [32, 33, 39]) or end-user eco-feedback tools (e.g. [7, 10–12, 25]). The former research tends to focus on the technical domain-specific challenges, where as the latter is more concerned with the suitability and effectiveness of such tools for 'nudging' [37] its users to continue using the tools and effecting positive behavioural changes to reduce the carbon footprint of its users. For a system like the EnerCoach tool, where domain experts and end-users meet, and the motivation for use lies in other factors than changing individual behavior, these studies are quite limited in their applicability to this project. Nonetheless, some important insights into the cognitive processes involved in understanding complex, energy-related issues can be gleamed from the literature. Framing the tasks of understanding the reports generated by the system in the same way that users of eco-feedback tools are trying to 'make sense' of the feedback they garner from the app helps tailor potential measures to improve transparency to the need of the end-user. Sense-making processes - the activity of gaining *"[...] a meaningful and functional representation of some aspects of the world."* [22, p. 1] - are based on an iterative cycle of *"framing, elaborating and reframing data"* (see Tellioğlu et al. [36, p. 2], citing Klein et al. [19]). As Wood et al. show in their study of sense-making processes of users of an eco-feedback tool for households, incorporating contextual information with the numeric and graphical results of such displays can help improve user's understanding, but the conclusions drawn from these processes can still differ from person to person [41]. This underscores the 'wicked' nature of the problem of transparency even more: while certain measures might improve transparency for given aspects of a system, the conclusions drawn by the target audience might be very different depending on their backgrounds and motivations.

3 METHODOLOGIES

Bridging the technical and social domains, the methodologies for this study combine multiple approaches to contribute to the unpacking of the socio-technical assemblage that makes up the EnerCoach system and generate a *thick description* [13] of the system and its stakeholders. The first phase of the project involved a series of semi-structured qualitative interviews with administrators, energy consultants and community users of the system (N=8). During these, the relevant stakeholder and user groups were identified and their views on and uses of the system, as well as their needs and requirements towards the transparency of the system, were recorded and analyzed following Mayring's approach to qualitative inquiry by coding the transcripts [24]. To augment these empirical data, two half-day training sessions that are regularly offered to future users of the system with 15 participants each were observed and analyzed to assess the differences in algorithmic literacy and user knowledge of the various user groups. From a technical standpoint, a complete code-level review of the underlying system, including the data collected and the algorithmic processing for the reports, was performed. Furthermore, usage statistics were collected in cooperation with the company developing and maintaining the system. Merging the results of applying these methodologies, a situated ethnography of this algorithmic system [31] details the various interactions of involved parties with the system, the system's technical aspects and functionalities, as well as the current problems and challenges arising from these specific socio-technical configurations.

To tackle one of the core challenges identified in this first phase - a lack of transparency and understanding for the algorithmic processes underlying the data collection and reporting functionalities - the second phase involved a participatory design workshop to develop potential interventions for both the technical and social practice aspects of the system. Participatory Design (PD) transcends other human-centred design approaches in its core, ethical stance that *"[...] people have the basic right to make decisions about how they do their work"* [29, p.65], and typically involves users not simply as sources of information (i.e. to ask them about their needs), but as active participants in the decision making process of design and development. In the case of EnerCoach, the use of PD methodologies facilitated two goals: on the one hand, the workshops allow participants to contribute to the decisions and to add their expertise on what measures help them understand the underlying processes. On the other hand, observing the sense-making activities happening throughout the workshop process itself provides further insights into what techniques facilitate a better understanding of the system.

The following section presents a *thick description* [13] of the EnerCoach tool, the context and practice of its use and the core issues of transparency within the system based on the analysis of the qualitative and quantitative data collected throughout the first phase of this project.

4 THE ENERCOACH ACCOUNTING TOOL

EnerCoach is a collaborative energy accounting tool currently used by over 600 Swiss communities to collect and visualize data about energy consumption for electricity and heating, as well as water consumption, of community-owned buildings. An example of the larger group of eco-feedback and management tools, the EnerCoach system enables communities to fulfill the requirements of sustainability initiatives and programs such as the European Energy Awards (EEA)[2], which - as part of their measures recommended in their catalogues - require an energy accounting strategy to monitor buildings and facilities owned by the community. The EnerCoach tool is funded by the Swiss member group of the EEA, EnergyCities, which awards communities with the *EnergieStadt* label for completing the requirements of the program in a 4-year evaluation cycle, and currently grants communities access to the tool free of charge. The online version of the tool was based on a legacy version of a tool with the same name that was originally implemented via a complex and opaque Microsoft Excel sheet and distributed to communities as templates to be filled out. The downsides of this old tool - particularly the difficulties of providing updates to hundreds of communities via sending out updated Excel sheets - led to the development of the current, online and collaborative tool. Another benefit of the online version over the legacy tool is its multi-lingual capability: the unique context of Switzerland's multi-lingual population, with the four prevalent languages German, French, Italian and Rhaeto-Romanic, necessitates a multi-lingual tool to allow full utilization by the different regions of Switzerland. In addition to English as a base language, the current tool implements three out of these four languages (German, French and Italian).

The stakeholders and users of the tool can be roughly grouped as follows: the *EnerCoach working group*, consisting of Swiss sustainability policy and energy accounting experts, serve as a steering committee and define goals for the further development of the tool in general and calculation policies in particular. Various *energy consulting and research companies* serve as point of contact for different regions in Switzerland, provide hotline support via phone and email to *end users* of the system, and mediate between both these users

[2]https://www.european-energy-award.org/welcome-to-the-european-energy-award

and *WIENFLUSS*[3], the company developing and maintaining the EnerCoach system. End users of the system take various roles as well: *community users* typically work within a single (or seldom more than one) community to enter data and generate reports, while *building managers* only focus on data entry for buildings assigned to them. Outside of the communities themselves, *energy consultants* hired by communities support these processes by validating the data entered and give suggestions for potential remediation measures to improve the sustainability performance of the community's buildings. Finally, *energy auditors* - themselves often *energy consultants* in other capacities - perform assessments of the communities performance and planned / implemented measures as part of the certification process of the EnergyCities / EEA programs. Figure 1 illustrates these users, groups and their relationships.

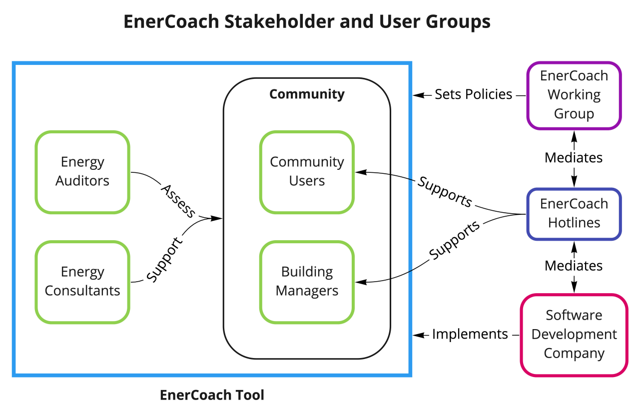

Figure 1: EnerCoach stakeholder groups and their interactions

4.1 Technical implementation

From a technical standpoint, the EnerCoach system implements a hierarchical data model as illustrated by the simplified diagram in figure 2. Communities have some inherent attributes - including addresses and geographical location, contact data and the nearest weather station used to incorporate climate data into the reports - and serve as containers for organisational units and building objects on the one side, and a set of energy mixes describing the shares of different energy carriers for user-defined time periods. Such an energy mix may, for instance, describe that a given community's electricity supply for the time period of January 1st until December 31st, 2020 would consist of 60% hydro power, 10% solar power and 30% fossil fuel based power, making the renewable share of energy production of this period 70%. For each of the 65 energy carriers available in the system, a number of factors are predefined that determine the energy carrier's primary energy factor, greenhouse gas emissions, share of renewable energy, density, heating value and unit.

Building objects themselves can house any number of electricity, heating or water meters, which are the primary point of data collection for energy consumption. Similar to the energy mix periods,

[3] see https://www.wienfluss.net

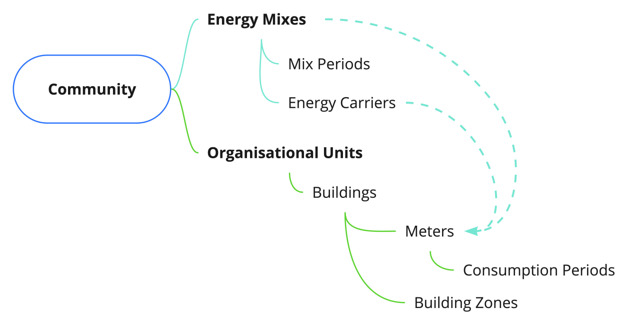

Figure 2: Simplified data model of a community in the EnerCoach energy accounting system.

consumption periods allow entering the energy consumed, energy produced (in the case of on-premise energy sources like photovoltaic power) and costs incurred for a given meter and time period defined by a start date and an end date. To allow detailed reporting on the sustainability impact of these consumptions, each meter is assigned either one of the mixes defined at the community level, or directly connected to an energy carrier such as local solar power.

To allow a comparison between the energy efficiency of different buildings, the *utilization reference area* of a building is required to calculate the energy expenditures per square meters (referred to as *key figures* throughout the system); these data are entered as different building zones within one building and classified as one of twelve building zone categories standardized by the Swiss Society of Engineers and Architects (SIA)[4] [1]. A community collecting data about a local school could, for instance, define an object with three zones - one for the school building itself, one for the gym, and one for the living quarters of a custodian living on the premises - all with their own SIA category, reference area in square meters and utilization factors for electricity, thermal and water consumption. These building categories play a vital role in the determination of energy efficiency and comparability of different building types: since the average energy requirements for heating a public pool are substantially larger than those of residential apartments, the target values for energy consumption per m^2 must reflect these differences, lest the public pool always gets classified as deficient compared to other building types.

A core aspect of the EnerCoach system are the reporting functions that aggregate the data entered and provide both descriptive and normative insights into the performance of the community. Currently, a total of 11 detailed reports, as well as one overview dashboard collecting all other other reports on one page, are available for either the entire community or single objects. These reports include basic aggregations of energy consumption, CO_2 emissions and incurred energy costs, time-series aggregating key figures (energy consumption in kWh per m^2), as well as aggregations calculating renewable energy shares and energy efficiency in various degrees of detail. Each report type also features a set of parameters, allowing more granular insights into the different energy / resource use (electricity, thermal and water consumption),

[4] orig. "Schweizerischer Ingerieur und Architektenverein (SIA)"

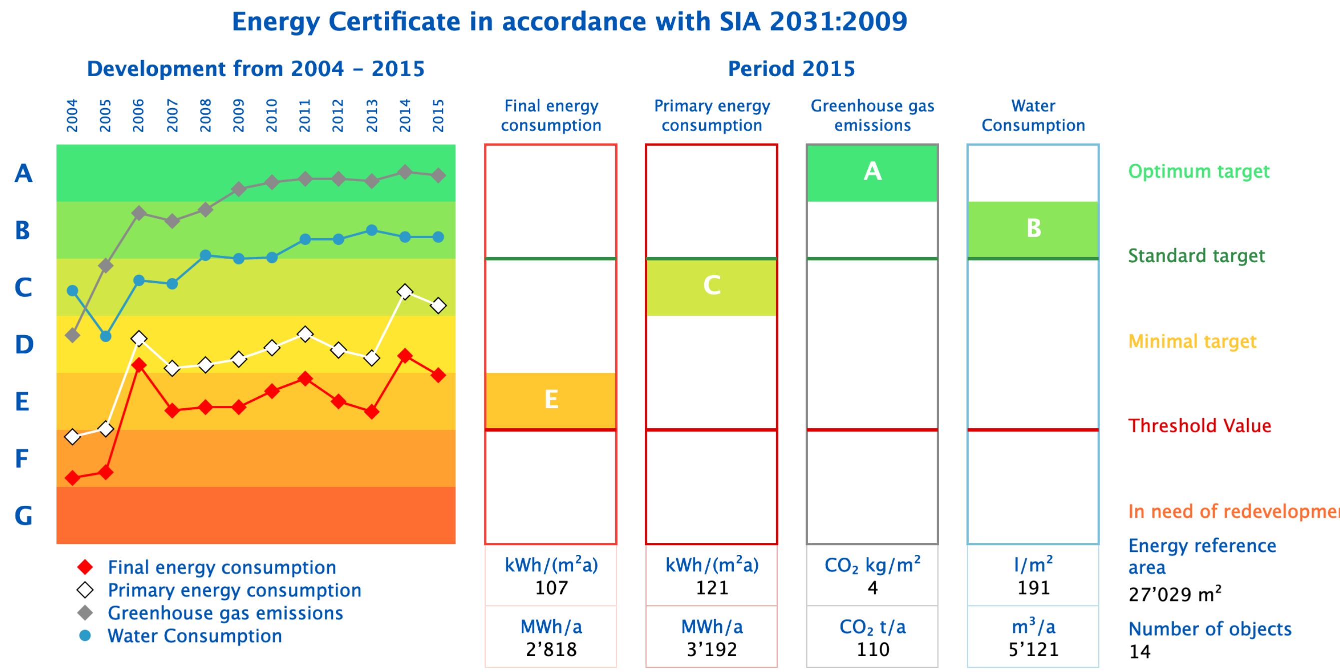

Figure 3: Sample energy certificate report.

specific time periods, and filtering the results by building categories or organisational units. The most relevant report for community auditing processes is the energy certificate report shown in figure 3, which directly evaluates the community performance against consumption and CO_2 emission target values based on building zone categories and reference areas, and delivers a rating of either the entire community or single objects on an 8-point scale from A to G. These target values depend on numerous factors, including the composition and categories and building zones, their size, as well as climate data adjustments to reflect that energy expenditures for heating in cold winters is expected to be higher.

Far from simple statistical aggregations, the reports represent a manifestation of various larger *policy decisions made durable* in the form of an algorithmic system, in the same sense that Latour described technology as *"society made durable"* [21]. As the scientific literature in the sustainability and climate change domains show, a multitude of approaches and competing standards exist for many of the relevant calculations, from CO_2-emission factor calculation to rating building efficiency [40]. Any practical implementation of these calculations must inevitably be the result of deliberations of trade-offs between accuracy and complexity as well as specific decisions that - in the case of EnerCoach - take on a normative character insofar as they influence the potential conclusions users draw from the generated results. An illustrative example of this is the calculation of the key figures for thermal production, i.e. energy used for heating. Here, the EnerCoach tool automatically introduces a factor of 2.0 to any energy consumed if and only if the heating system used is direct electrical heating. As was revealed by one of the interviewees responsible for this decision, the reason for this choice was the low energy conversion efficiency of electrical heating when including the power generation and transfer loss when using power from fossil fuel power plants, and reflects the national energy policies meant to nudge end users towards phasing out direct electrical heating in favor of other heating methods that have a better energy conversion efficiency.

Beyond such specific policy-based exceptions to the default calculations employed in the system, another factor contributing to the complexity was the decision to allow arbitrary time periods for both energy mixes and consumption periods. The original Microsoft-Excel based EnerCoach tool allowed data entry only for yearly periods, which was a notable source of errors: since periodic energy bills listing the energy mix used for the electricity supplied to the community often differed in their time frames from the time periods that meters were read in, communities needed to manually average their yearly consumptions per meter to match the energy mix periods. The current implementation standardizes this process by treating all consumption and mix data as daily values, and thus automates the process of normalizing time periods for these calculations, but subsequently also increases the complexity of the calculations.

4.2 Function-specific transparency requirements

Based on the interview results, the main functions of the EnerCoach tool can be categorized as follows: (1) Allowing communities to collect data about community-owned buildings, building zones, meters, energy mixes and continuous energy consumption, (2) visualizing the collected data in the form of reports and dashboards and (3) providing energy consultants and auditors of the EnergyCities / EEA programs access to the cities performance related to energy expenditure and sustainability in order to complete the certification process mandated by the EEA and similar programs. These three functions necessitate the coordination of a large and diverse group of users and stakeholders: energy and sustainability experts, city officials and administrators and local facility / building managers all utilize the tool in a different way, and have different requirements towards the transparency and explainability of the system. While local facility managers may only be responsible for entering consumption data of a single building, a city's administrative personnel would use both data entry and reporting functions and maintain data about the supply side of the energy expenditure (for instance, the composition of energy mixes based on different energy carriers such a wind energy, coal or nuclear energy). Furthermore, energy auditors and consultants will be utilizing the reporting functions first and foremost to draw conclusions as well as propose and assess measures based on the interpretation of those results. Community sizes ranging from small communities with less than 200 inhabitants to larger cities with over 100.000 inhabitants require very different collaboration efforts of involved parties as well: while a small town may well be able to maintain use of the tool through a single responsible person, larger communities require teams with varied responsibilities and skills collaborating in their use of the tool. Furthermore, policy makers such as the EnerCoach working group or the SIA impact the tool through changes to standards such as the SIA Norm 380 [23] detailing calculations of energy target and threshold values for different building categories.

All of these three main functions of the tool - *data collection*, *data aggregation and visualization*, and *auditing*, present unique requirements and challenges for the transparency of the system. In terms of *data collection*, multiple users can contribute to the data pool of a community by adding consumption data for meters, editing properties of building objects and zones, or adjusting the electricity, gas or district heating mixes that feed into the meters. With any of these actions, they can impact other user's experience and results of the tool significantly: changing the energy carrier mix of a community would impact all meters utilizing that mix, and changing the utilization area of a single building zone can change the energy certificate grades of the entire community. In this sense, transparency (or a lack thereof) mostly meant ex-post *traceability* of other user's actions. As one energy consultant explained, the audit process often calls into question the validity of the data, and would necessitate tracking down the user responsible for that data entry - a process that is not currently supported by the tool itself and was thus relegated to the social realm.

For the functionalities providing *data aggregation and visualization*, transparency pertains to two variants of explainability: *transparency of the model* and *post-hoc explanation* as described by Mittelstadt et al. [26]. The former refers to the general understanding users have about the internal model of the system, the calculations involved and the complex interplay between the different entities and concepts manifested in the system, whereas the latter represents the system's capability to support sense-making processes after the system has finished its calculations and is presenting a result in the form of a report. The EnerCoach tool mainly addresses these two aspects through multiple non-technical measures and processes. First, training sessions for potential users offered by the providers of the system aim to not only teach future users how to use the system, but also educate them on the basics of energy accounting and reporting. While a rather rudimentary online documentation for the system exists, the quantitative data analysis of user logs showed that it is rarely used, leaving the training sessions as primary point of learning the skills necessary for the basic use of the system. Second, a support hotline offers help to users for specific questions, and often involves mediating between users and the system's developers to verify specific calculation results. Third, the collaborative nature of the tool allows for energy consultants with a broader understanding of the underlying calculations and policies to support end users in correcting errors in the *data entry* process leading to incorrect results, and also serves as a fail safe for potential implementation errors or bugs: for instance, energy consultants or auditors discovered a number of edge cases for specific constellations of meters, zones and mixes that yielded wrong results and communicated those to the developers to address.

Lastly, transparency for the *auditing* functionality of the system is mostly located in the *social* aspect of the socio-technical assemblage that makes up the EnerCoach tool as well. As energy auditors access the system to shepherd communities through various certification processes, they verify the correctness of the data entered and require transparent access to the data sources (such as the energy bills listing the energy carrier shares making up the energy mixes supplied to the buildings). Collaborating with the community users, they trace errors in the data entry and provide transparent feedback on potential sustainability measures to improve the community's performance, but they do so mostly outside of the system.

4.3 Transparency deficiencies

Out of the various types of transparency required for the functionalities outlined above, the issues of *model transparency* and *post-hoc explanations* [26] were identified as most deficient by the participants of the study. These deficiencies relate to both the end-users (as communicated to the hotline) and the experts staffing the hotline themselves, as well as the energy consultants and auditors utilizing the tool for their own purposes. The focus on these two issues is an interesting result insofar as they are not directly related to human actants in the system (like the issue of *traceability* regarding other user's actions) but rather just to the system itself. While traceability as a deficiency was mentioned, the social remediation strategies (i.e. tracing other user's actions by social means rather than technical ones) seem to be working well enough that it was not seen as a pressing issue.

In terms of model transparency, some of the core relationships between different data entities (building objects, meters, mixes) were not clear in the context of report generation. As one energy consultant explained, the number of factors that influence target and threshold values for consumptions of buildings based on their zone definitions is particularly opaque - and requires the users trust that the values presented are correct. Since these values also depend on user-entered data, such as the zone's reference area or the zone's utilization factors for electricity, heating and water, verifying the plausibility of the results can be a hassle. Furthermore, the fact that different target and threshold value calculation methods are used for different reports eluded most of the interview participants.

As a final critique towards model transparency, the order of calculation steps was unclear to both experts and end users in many cases. Since it makes a difference for the aggregations of the key figure report of a community whether all consumptions and reference areas are summed up first and then divided by each other, or whether single key figures of buildings are calculated first and then aggregated through weighted averages, these implementation details - while highly technical - would be a crucial information to assist in troubleshooting certain aberrant report results. As the hotline staffers and energy experts suggested, the level of understanding required to distinguish these two approaches most likely prohibits end users from grasping the difference, making the expert users primary targets for transparency measures aimed at elucidating these details.

In terms of *post-hoc explainability*, a number of deficiencies were noted as well. Since the report generation for larger communities can include aggregating hundreds of objects and thousands of consumption periods, it can be extremely difficult to trace the one object or meter with missing or wrong data that yielded an implausible report result. The complexity of normalizing the time frames between mixes, consumption periods and building zones also contributes to these issues; for instance, it is often unclear to the users whether or not the reference area of a building contributes to the communities total key figures if the building in question has no consumption data entered for a given time frame, thus skewing the community key figure reports towards a more positive result. Another issue that became apparent through the qualitative interview results was the lack of understanding for the internal plausibility checks involved in calculating the report. Due to mathematical necessities, certain calculations can not be performed if there are missing data; for instance, no key figures can be calculated if the values for reference areas are missing. The current implementation thus includes automatic removal of certain buildings and meters from the reports if the available data prohibits the calculation - a fact that was not clear to the users despite the fact that each report presents a list of missing data points underneath the final visualization to simplify tracking and correcting these issues.

Finally, the interview analysis underscored the interdependence of *model transparency* and *post-hoc explainability*. On the one hand, any measures to better explain the results require a certain level of understanding about the underlying models and calculations; on the other hand, measures to improve the understanding of the underlying models remain difficult to grasp and somewhat irrelevant without concrete examples and use-cases. As one energy-expert described it, certain features of the reports are rarely relevant for his work, and he subsequently was neither interested in their underlying calculations nor planning on using them in the future.

5 PARTICIPATORY DESIGN WORKSHOP

To address the issues regarding transparency identified in section 4.3, a participatory design workshop with a group of five hotline staffers, members of the EnerCoach working group and energy experts/consultants was held to explore and co-design measures to improve both *model transparency* and *post-hoc explanations*. The workshop participants had in-depth knowledge of the needs of the end-users of the system through years of experience staffing the hotline as well as contributing to the EnerCoach working group decisions that shape the specific implementations of the EnerCoach reports. Furthermore, four of the particpants also had experience as energy auditors, using the system themselves to support communities in their efforts to increase the sustainability of their communal energy consumption.

The first issue addressed was how to visualize one of the more complex, but highly important reports: the energy certificate report shown in figure 3. Based on the suggestion of one of the interview participants to start with a *"common visual language"*, model elements such as building objects, meters, mixes and zones, as well as calculated entities such as target and threshold values, were created as pre-printed cards with a symbolic representation of the entity. Figure 4 shows a subset of these card designs.

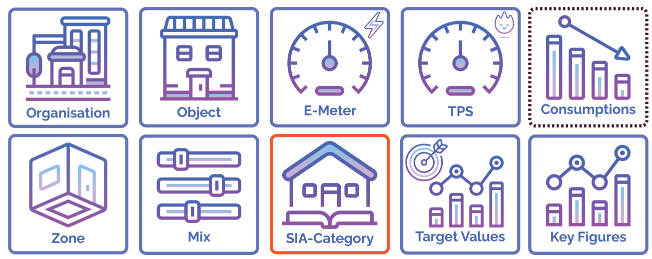

Figure 4: Selection of EnerCoach entity cards used in the participatory design workshop.

The participants were then asked to try to visualize the process necessary to calculate the energy certificate report to the best of their understanding, using the cards, pen and paper, and to narrate the report generation based on this visualisation. After a brief discussion and some clarifications on the details of the task, the group chose to assemble a type of flowchart reminiscent of algorithm visualizations used in teaching computer science students as surveyed by Hundhausen et al. [17]. Figure 5 shows some impressions of this cooperative group task in action.

This process also led to some lively discussions about the actual calculations and processes as understood by the participants, and shed further light on the particularly opaque aspects of the system. One interesting aspect that became particularly clear during this step was the confusion about the source of the data used for different entities. As one participant noted, the report includes both static data points entered by the user only once and changed very infrequently (such as building information or building zone data), data that need to be entered annually or more frequently (such as mix data or consumption data), as well as system-wide data points unchangeable by the user (such as energy carrier factors, SIA categories or climate data). The difference between these three types of data was shown to be highly relevant in the workshop, since limiting the search for faulty data to only user-entered data that needed frequent, yearly updates would most often suffice to resolve issues with implausible results. Based on a suggestion by the participants, the entity cards where subsequently adapted to reflect these differences through their border frame: blue for static, user-entered data, dashed blue for annually entered data and orange for system-wide data (the illustration in figure 4 already incorporates this change).

Figure 5: PD workshop visualization process.

When the group was satisfied with their visualization, a comparison with a visualization of the same report, but created by the developers of the system from a more technical standpoint, was discussed, focusing on the different interpretations of the processes involved. This comparison explicated the different perspectives the system developers and (expert) users had on the calculation process, and helped highlight the necessary balance between technical accuracy and comprehensibility of such a visualization. Discussing the usefulness of such visualizations for EnerCoach stakeholders, the group agreed that it would be particularly helpful for hotline staffers and energy consultants to be able to refer back to this resource when dealing with user questions. Distributing the graphics directly to the end user was seen as less promising, since the requirements for model transparency from the point of end-users were not as strict and could better be covered by the hotline.

The second part of the PD workshop focused on concrete measures that could be implemented within the EnerCoach system to support better *post-hoc explainability* of results. The participants used laminated printouts of reports and other system interfaces to create mockups of potential features that would help trace results and support the sense-making processes involved in reading the reports. The suggestion deemed most promising by the participants addressed the disconnect between user-entered data and the results of the reports, making it difficult to differentiate anomalous results, such as years with particularly low key figures, from data entry errors. The group proposed displaying user-defined annotations or comments next to the report, which could be entered alongside the meters and consumption data to explain specific data points. To give an example, a municipal building undergoing renovation for the better part of a year might have had very little energy consumption for heating if the building was not used during the winter months. In the key figure report, that year would show up with a significant drop in energy consumption per square meter; adding a comment when entering consumption data for the year explaining the anomaly (e.g. "Renovation: January - June, 2020") and displaying these comments next to the report output would immediately clarify this seemingly anomalous report result. This solution proposed by the participants also correlates with the results of the study by Wood et al. [41], which suggested that contextual information can support sense-making processes of eco-feedback tools.

Further suggestions focused on interface adaptations to clarify missing data, providing context-specific help texts that would explain certain features of the reports, and unifying terminology throughout the system and reports in all supported languages.

5.1 Results and implementation

Finalizing the workshop, the participants assembled a priority list of measures they deemed most promising that would be implemented and tested in the EnerCoach system. Following the positive feedback for flowchart visualizations, one priority was the design of two such graphics for the energy certificate and energy consumption reports. Figure 6 shows the resulting flowcharts, which were shared with the workshop participants and hotline staffers. Feedback gathered after distributing the final charts was generally positive: the charts were seen as helpful in gaining a general system overview and being reminded of the constituent parts of the specific reports. For instance, listing the various energy carriers aggregated into categories for the energy consumption report was regarded as helpful information, since these details were otherwise hidden within the original specifications of the system and not otherwise accessible. Furthermore, the approach of defining a common visual language was seen as a beneficial to the sense-making processes involved in understanding the report calculations. As described in section 2, this relates to the iterative nature of sense-making processes: As users try to understand the results of a report, they may need to go back and forth between data entry forms and report results to correlate the two, and shared visual representations help recognize connected entities such as building zones or meters. This observation also aligns with the results of decades of HCI and cognitive science research into the use and benefits of pictorial representations such as easily recognizable visual metaphors (e.g., a house icon for the building objects in this case) [4, p.225-227].

The second suggested feature that was implemented was to display contextual information in the form of comments next to the key figure reports. While it was not yet possible to qualitatively evaluate this feature, quantitative analysis of the user logs and database show that 23.48% of thermal production systems, 9.20% of electricity meters and 6.44% of water meters now have comments attached, suggesting - together with some feedback gathered from the hotline staffers - that the feature has at least found some adoption by and is useful to users.

6 FINDINGS

The results of the participatory design workshops and subsequently implemented features allow two observations: one concerns the applicability of participatory design itself towards algorithmic transparency, and the other relates to the results themselves.

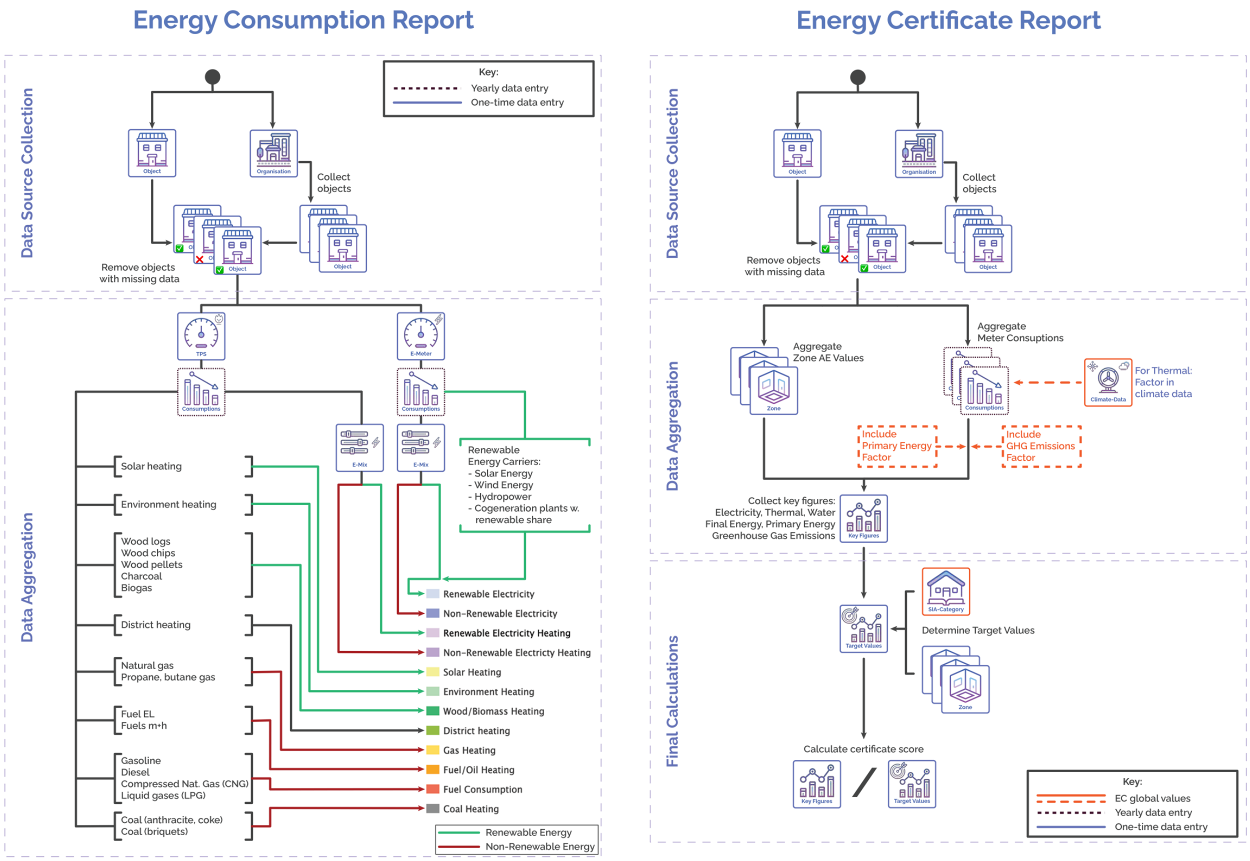

Figure 6: Final report visualizations for energy certificate and energy consumption reports.

Methodologically, the participatory design workshop proved to be a highly appropriate tool to engage expert users with pre-existing knowledge about the system, and helped both highlight existing deficiencies in the system's transparency and constructive, practical work on remedies for the issue. Participants contributed materially and conceptually to the solutions, and were engaged beyond the workshop itself in the process of implementing and evaluating the solutions. Furthermore, the workshop helped bridge the gap between the system developers and expert users, and thus contributed in and of itself to the participant's understanding of the EnerCoach system. This also resulted in an unexpected side-effect: by gaining an understanding for the complexities of the system and its calculations, the participants showed a remarkable restraint in requesting unrealistic or sweeping changes to the system, an issue well known in the software development community as *feature creep*. Instead, they focused on reasonable solutions fitting well within the established technical and interface frameworks. Finally, the participant's contributions to design and implementation decisions helped invest them further in the process and heightened their agency: given the relationship of accountability and transparency as outlined by Kemper et al., the workshops contributed to elevating the participants to become a more *"critical audience"* of the algorithmic system [18].

Concerning the measures themselves (including the report visualizations and implemented features), further evaluation of their effectiveness for supporting transparency in its various forms will be needed. Nonetheless, the first feedback gathered and preliminary evaluations already suggest a significant potential of even comparably small changes to the interface, such as the display of contextual information to explain report results, for improving *post-hoc explainability* of the system. The algorithm visualizations themselves also garnered promising feedback, and showed the importance of targeting specific measures for *model transparency* towards appropriate audiences. This can be seen as an important guideline for future developments, particularly in algorithmic systems with diverse user groups and stakeholders with varying degrees of domain knowledge and algorithmic literacy.

6.1 Limitations and outlook

While the study results in and of themselves are promising, the process has also underscored the 'wickedness' of the problem of algorithmic transparency once more. As the participants of the workshop also pointed out, no one-size-fits-all solution for algorithmic transparency can be possible, not even within a single system like EnerCoach. Even evaluating the measures developed by the group is proving to be a difficult task and will require further work.

To fully grasp the impact of small changes to such a large system, a mixed-method approach including both more in-depth qualitative and quantitative evaluations will be needed; borrowing methodologies from HCI and user-interface design (such as A/B testing) might prove successful in this endeavor. Future work will also include an expansion of the use of algorithm visualization; prior work (e.g. by Hundhausen et al., [17]) suggests alternate modalities of these visualizations such as interactive visualizations implemented directly in the system might have the potential to lower the cognitive effort of sense-making.

Finally, the project's participants were limited to more advanced users. Even though the inclusion of hotline staffers guaranteed some insights into the needs and challenges to community end-users, targeting a study specifically towards these end-user groups would be a worthwhile approach to gain further insights on the issues of transparency from their point of view.

7 CONCLUSIO

In this study, I presented a novel approach to the 'wicked problem' of algorithmic transparency through participatory design for the case study of the EnerCoach collaborative energy accounting system. Through a combination of qualitative and quantitative methods, a situated ethnography of this algorithmic system conceptualized *"as culture"* [31] served to highlight the complexity of the issue and detailed various challenges to transparency categorized by the different functionalities of the EnerCoach system. *Model transparency* and *post-hoc explanations* [26] were shown to be the biggest challenges as identified by users and stakeholders of the system: the former refers to the user's understanding of the internal processes and calculations of the system as a whole, the latter describes the user's ability to make sense of the system's outputs in the form of energy-related reports and aggregations. Utilizing participatory design as a methodology helped engage expert users in co-creating measures to improve the situation and shed further light on these difficult issues. The chosen methodology also provided valuable insights into the different perspectives taken by the various user groups and bridged the gap between the stakeholders of the system, namely system developers and expert users. The visualizations and features resulting from the participatory design workshops also show a promising potential to support the sense-making processes necessary to understand the system better and troubleshoot implausible results, thus aiding the hotline staffers and expert users in their work within the system. In a more general sense, the study also provides insights into the contextual nature of algorithmic transparency: different interpretations of transparency are shown to be dependent on the functionalities of the system, and both the ethnographic work and the participatory design workshop also show the important role social and procedural (as opposed to purely technical) measures can play in enabling transparency and, subsequently, accountability of algorithmic systems.

While further work in evaluating the results will be needed, the project already represents significant improvements to the EnerCoach system and the first step in the incremental process of reaching 'satisficing' solutions to the 'wicked problem' of algorithmic transparency.

ACKNOWLEDGMENTS

The various stakeholders of the EnerCoach system who so graciously volunteered to take part in the PD workshops and interviews made this study possible and deserve our gratitude. Furthermore, the project management team at WIENFLUSS deserves recognition for their eargerness to cooperate with this study, supporting the code review and offering their insights into the EnerCoach community.

REFERENCES

[1] 2021. sia | schweizerischer ingenieur- und architektenverein. https://www.sia.ch/de [Online; accessed 12. Feb. 2021].
[2] Doris Allhutter, Florian Cech, Fabian Fischer, Gabriel Grill, and Astrid Mager. 2020. Algorithmic profiling of job seekers in Austria: how austerity politics are made effective. *Frontiers in Big Data* (01 2020), 1 – 28. https://doi.org/10.3389/fdata.2020.00005
[3] M Ananny and K Crawford. 2016. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* (2016). https://doi.org/10.1177/1461444816676645
[4] Alan F. Blackwell. 2001. Pictorial Representation and Metaphor in Visual Language Design. *Journal of Visual Languages & Computing* 12, 3 (2001), 223–252. https://doi.org/10.1006/jvlc.2001.0207
[5] Robert Brauneis and Ellen P Goodman. 2017. Algorithmic Transparency for the Smart City. *SSRN Electronic Journal* (2017). https://doi.org/10.2139/ssrn.3012499
[6] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. 3, 1 (01 2016), 205395171562251 – 12. https://doi.org/10.1177/2053951715622512
[7] Florian Cech and Marlene Wagner. 2019. Erollin' on green: A case study on eco-feedback tools for emobility *(Communities and Technologies 2019)*. 121 – 125. https://doi.org/10.1145/3328320.3328402
[8] Nicholas Diakopoulos and Michael Koliska. 2016. Algorithmic Transparency in the News Media. *Digital Journalism* 5, 7 (08 2016), 809 – 828. https://doi.org/10.1080/21670811.2016.1208053
[9] Geraldine Fitzpatrick. 2003. *The Locales Framework, Understanding and Designing for Wicked Problems*. https://doi.org/10.1007/978-94-017-0363-5
[10] Myriam Frejus and Dominique Martini. 2015. Taking into Account User Appropriation and Development to Design Energy Consumption Feedback *(the 33rd Annual ACM Conference Extended Abstracts)*. 2193 – 2198. https://doi.org/10.1145/2702613.2732718
[11] Jon Froehlich, Leah Findlater, and James Landay. 2010. *The design of eco-feedback technology*. ACM. https://doi.org/10.1145/1753326.1753629
[12] Silvia Gabrielli, Paula Forbes, Antti Jylhä, Simon Wells, Miika Sirén, Samuli Hemminki, Petteri Nurmi, Rosa Maimone, Judith Masthoff, and Giulio Jacucci. 2014. Design challenges in motivating change for sustainable urban mobility. *Computers in Human Behavior* 41, C (12 2014), 416 – 423. https://doi.org/10.1016/j.chb.2014.05.026
[13] Clifford Geertz. 1973. *The Interpretation of Cultures*. Basic Books, Inc. https://chairoflogicphiloscult.files.wordpress.com/2013/02/clifford-geertz-the-interpretation-of-cultures.pdf
[14] R Stuart Geiger. 2017. Beyond opening up the black box: Investigating the role of algorithmic systems in Wikipedian organizational culture. *Big Data & Society* 4, 2 (07 2017), 205395171773073 – 14. https://doi.org/10.1177/2053951717730735
[15] Susse Georg and Lise Justesen. 2017. Counting to zero: accounting for a green building. *Accounting, Auditing & Accountability Journal* Volume 30, Issue 5 (2017), 1065–1081. https://doi.org/10.1108/aaaj-04-2013-1320
[16] Tarleton Gillespie and Nick Seaver. 2016. Critical Algorithm Studies: A Reading List. https://socialmediacollective.org/reading-lists/critical-algorithm-studies/
[17] Christopher D. Hundhausen, Sarah A. Douglas, and John T. Stasko. 2002. A Meta-Study of Algorithm Visualization Effectiveness. *Journal of Visual Languages & Computing* 13, 3 (2002), 259–290. https://doi.org/10.1006/jvlc.2002.0237
[18] Jakko Kemper and Daan Kolkman. 2018. Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society* 0, 0 (06 2018), 1 – 16. https://doi.org/10.1080/1369118x.2018.1477967
[19] Gary Klein, Brian M Moon, and Robert R Hoffman. 2006. Making Sense of Sensemaking 1 - Alternative Perspectives. *IEEE Intelligent Systems* 21, 4 (2006), 70 – 73. https://doi.org/10.1109/mis.2006.75
[20] Alexander Kunze, Stephen J. Summerskill, Russell Marshall, and Ashleigh J. Filtness. 2019. Automation transparency: implications of uncertainty communication for human-automation interaction and interfaces. *Ergonomics* 62, 3 (2019), 1–22. https://doi.org/10.1080/00140139.2018.1547842
[21] Bruno Latour. 1990. Technology is Society Made Durable. *The Sociological Review* 38, 1_suppl (1990), 103–131. https://doi.org/10.1111/j.1467-954x.1990.tb03350.x
[22] Christian Lebiere, Peter Pirolli, Robert Thomson, Jaehyon Paik, Matthew Rutledge-Taylor, James Staszewski, and John R Anderson. 2013. A functional

model of sensemaking in a neurocognitive architecture. *Computational Intelligence and Neuroscience* 2013, 4124 (2013), 921695 – 29. https://doi.org/10.1155/2013/921695
[23] Martin Lenzlinger. 2021. SIA Norm 380. https://www.sia.ch/de/themen/energie/artikelbeitraege/detail/article/energetische-berechnungen-neu-geordnet [Online; accessed 12. Feb. 2021].
[24] Philipp Mayring. 2002. *Einführung in die qualitative Sozialforschung* (beltz studium ed.). Beltz Verlag.
[25] Johanna Meurer, Dennis Lawo, Lukas Janßen, and Volker Wulf. 2016. Designing Mobility Eco-Feedback for Elderly Users *(the 2016 CHI Conference Extended Abstracts)*. 921 – 926. https://doi.org/10.1145/2851581.2851599
[26] Brent Mittelstadt, Chris Russell, and Sandra Wachter. 2019. Explaining Explanations in AI *(FAT* '19)*. 279 – 288. https://doi.org/10.1145/3287560.3287574
[27] F Pasquale. 2015. *The black box society: The secret algorithms that control money and information*. Harvard University Press. https://doi.org/10.4159/harvard.9780674736061
[28] Horst W. J. Rittel and Melvin M. Webber. 1973. Dilemmas in a general theory of planning. *Policy Sciences* 4, 2 (1973), 155–169. https://doi.org/10.1007/bf01405730
[29] Toni Robertson and Ina Wagner. 2013. Ethics: engagement, representation and politics-in-action. In *Routledge International Handbook of Participatory Design*. 84–105. https://doi.org/10.4324/9780203108543-11
[30] Nick Seaver. 2013. Knowing algorithms. *Media in Transition* 8 (2013).
[31] Nick Seaver. 2017. Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4, 2 (07 2017), 205395171773810 – 12. https://doi.org/10.1177/2053951717738104
[32] Ling Shao, G.Q. Chen, Z.M. Chen, Shan Guo, M.Y. Han, Bo Zhang, T. Hayat, A. Alsaedi, and B. Ahmad. 2014. Systems accounting for energy consumption and carbon emission by building. *Communications in Nonlinear Science and Numerical Simulation* 19, 6 (2014), 1859–1873. https://doi.org/10.1016/j.cnsns.2013.10.003
[33] Pratima Singh and Arun Kansal. 2018. Energy and GHG accounting for wastewater infrastructure. *Resources, Conservation and Recycling* 128 (2018), 499–507. https://doi.org/10.1016/j.resconrec.2016.07.014
[34] Nicola J. Spalding and Terry Phillips. 2007. Exploring the Use of Vignettes: From Validity to Trustworthiness. *Qualitative Health Research* 17, 7 (2007), 954–962. https://doi.org/10.1177/1049732307306187
[35] Yolande Strengers. 2011. Beyond demand management: co-managing energy and water practices with Australian households. *Policy Studies* 32, 1 (Jan. 2011), 35–58. https://doi.org/10.1080/01442872.2010.526413
[36] Hilda Tellioglu, Michael Habiger, and Florian Cech. 2017. Infrastructures for Sense Making. (2017).
[37] Richard H. Thaler and Cass R. Sunstein. 2009. *Nudge: improving decisions about health, wealth, and happiness*. Choice Reviews Online, Vol. 46. https://doi.org/10.5860/choice.46-0977
[38] Matteo Turilli and Luciano Floridi. 2009. The ethics of information transparency. *Ethics and Information Technology* 11, 2 (03 2009), 105 – 112. https://doi.org/10.1007/s10676-009-9187-9
[39] Jan Warnken, Melanie Bradley, and Chris Guilding. 2004. Exploring methods and practicalities of conducting sector-wide energy consumption accounting in the tourist accommodation industry. *Ecological Economics* 48, 1 (2004), 125–141. https://doi.org/10.1016/j.ecolecon.2003.08.007
[40] Thomas Wiedmann and Jan Christoph Minx. 2008. A Definition of 'Carbon Footprint'. In *Ecological Economics Research Trends*, Carolyn C. Pertsova (Ed.). Nova Science Publishers, Hauppauge NY, USA, 1–11.
[41] Georgina Wood, Rosie Day, Emily Creamer, Dan van der Horst, Atif Hussain, Shuli Liu, Ashish Shukla, Obiajulu Iweka, Mark Gaterell, Panagiotis Petridis, Nicholas Adams, and Victoria Brown. 2019. Sensors, sense-making and sensitivities: UK household experiences with a feedback display on energy consumption and indoor environmental conditions. *Energy Research & Social Science* 55 (2019), 93–105. https://doi.org/10.1016/j.erss.2019.04.013