Researching AI Legibility through Design

Joseph Lindley, Haider Ali Akmal, Fraziska Pilling, Paul Coulton
Imagination
Lancaster University
{j.lindley, h.a.akmal, f.pilling, p.coulton}@lancaster.ac.uk

ABSTRACT
Everyday interactions with computers are increasingly likely to involve elements of Artificial Intelligence (AI). Encompassing a broad spectrum of technologies and applications, AI poses many challenges for HCI and design. One such challenge is the need to make AI's role in a given system legible to the user in a meaningful way. In this paper we employ a Research through Design (RtD) approach to explore how this might be achieved. Building on contemporary concerns and a thorough exploration of related research, our RtD process reflects on designing imagery intended to help increase AI legibility for users. The paper makes three contributions. First, we thoroughly explore prior research in order to critically unpack the AI legibility problem space. Second, we respond with design proposals whose aim is to enhance the legibility, to users, of systems using AI. Third, we explore the role of design-led enquiry as a tool for critically exploring the intersection between HCI and AI research.

Author Keywords
Artificial Intelligence; Machine Learning; Legibility, Human-Data Interaction, Research through Design.

CSS Concepts
• **Human-centered computing~HCI design and evaluation methods**

INTRODUCTION

While AI's role in our zeitgeist reflects the computing revolution of the late 20th and early 21st centuries, the philosophical complexities of creating artificial life have fascinated humanity for millennia. This ongoing interest is evident within the significant contrast between the mythology of stories *about* AI (e.g. *Talos, HAL9000,* and *I, Robot*), and the actual implementations of computer systems which *utilize* AI (e.g. *Google Search*, *Amazon Echo, FaceApp*). This distinction is manifested in how AI is commonly imagined by the general public as machines which exhibit qualities we associate solely with humanity as opposed to the *reality* of current implementations of AI—these existing systems compute answers using algorithms, which, at first glance, is a significantly more straightforward proposition than truly sentiment computers. These apparently distinct spaces, in fact, have shared ancestry. When Alan Turing posed the question 'Can machines think?' [52] he kickstarted the conflation of the philosophical and technical territories that we know of as AI today. Turing had noted that humans use available data, combined with logical reasoning, in order to learn. This was something which he surmised might also be achievable using a computer.

Whilst any concrete answer to Turing's question has remained aloof, innovations stemming from research attempting to answer this question have given rise to a plethora of computing techniques (e.g. expert systems, classifiers, machine learning, neural networks). Driven by the availability of cheap/powerful computing and increasingly abundant data, decades of sedate progress in AI have given way to rapid acceleration specifically around the use of neural networks, machine learning, and deep learning. Such progress—famously exhibited by Google DeepMind's game-playing AIs, which develops human-like strategies [22]—simultaneously drives the utilization of AI in a wide range of applications, but also gives credence to speculation about machines which really *can* think. Nonetheless the teachable, usable, and already-ubiquitous computing techniques (which we refer to as AI) and the 'grand challenge' of creating a machine that can think like a person, be emotional, and act creatively (which we *also* call AI), are quite different propositions and present different categories of problem for researchers and HCI scholars to consider. This apparently straightforward problem of conflated rhetoric is not to be underestimated. The AI field is more mature than was imaginable in the 1950s, yet Turing's legacy helps perpetuate flawed anthropocentric thinking [28]; it accelerates AI development in pursuit of a strategic advantage (but when driven by grandiose visions this leads toward palpable societal risks [10,23]); and increases anxiety about the role of AI arguably stifling its potential benefits [34]. While the ambiguity within the term AI is a complex matter, in this paper we adopt a relatively straightforward position. Whenever 'AI' is used to describe a system or innovation its 'legibility'—as the term is used in the emerging field of Human-Data Interaction (HDI)—is significantly reduced.

As a field of research which intersects with HCI, HDI is becoming established [11,27,44,48]. It reflects that the data produced through the use of computers, our interactions

transcend the devices themselves and have much broader impact. At its core HDI proposes exploring our relationships with data through the lenses of *legibility* (e.g. clarity about data use), *agency* (e.g. choice about processing data), and *negotiability* (e.g. how we trade off data access for functionality). While these concepts are distinct, they are interrelated and we can say that legibility is a necessary precursor to agency, and multiple manifestations of agency (or lack thereof) constitute scales of negotiability. As all implementations of AI utilize data (e.g. using data to train an AI model or processing data with an AI) the lenses of HDI provide a useful means through which to consider the socio-technical implications of AI. Given that describing an AI system's legibility would be achieved by designing legibility-enabling interactions, there is a significant nexus at the point HDI, HCI and design interact with one another.

Solving such a problem is a design challenge which by adopting RtD also presents a powerful and relevant context for research. As a maturing approach, debate continues about RtD's epistemology and methodology[1]. Whilst acknowledging these discussions, entering into them in detail is beyond the scope of the paper. This research employs RtD for two reasons. First, prior research into AI is multidisciplinary and made up of disparate elements that RtD's generative processes may meaningfully coalesce into artifacts which through their potential for utility and novelty are also artifacts of research. Second, AI legibility is as much a design challenge as it is a technical or theoretical one—it demands an array of original responses to the new challenges introduced by AI innovation—and therefore exploring legibility through design is a crucial constituent of the broader research strategy.

Signs and symbols pervade the modern world, conveying information to meet a wide range of needs. Though academic nuances describe each of these ideas in more detail [42,49], an obvious unifying factor is that they are visual tools used to articulate ideas. For example, used in a motoring context, signs provide warnings and instructions to motorists (e.g. no entry, stop, one-way). Symbols are more likely to be abstract, often carrying meaning by convention (e.g. chemicals, electricity, or male/female), they are often used as constituent parts on a sign to build up the overall meaning. Sometimes signs and symbols are combined together and then incorporated within or attached to objects or products, these 'labels' serve a wide variety of purposes (e.g. nutrition labels, marketing copy, or usage instructions). In the computing domain icons appear in user interfaces to help users in identifying the files, folders, and other features that they need to use. Increasingly icons from the computer world can be seen in our physical environment (e.g. a WiFi symbol to indicate access is available in a given location and relevant service logos such as Yelp, TripAdvisor or Trustpilot). Though sometimes purely used as a marketing device, logos are another type of sign that can symbolically represent that endorsement or accreditation from a particular organization (e.g. Fairtrade). Signs, symbols, logos, labels and icons make interactions more legible; highlighting where services are available, indicating how to use them, what the consequences of specific interactions might be, and communicating relevant information (e.g. about security, accreditation, and compatibility). In this paper we describe the process of researching and designing a visual language which uses cues from existing signs, symbols, icons and labels, to describe how products and services may utilize AI, and thus to explore how to enhance the legibility of AIs.

To this end the paper proceeds with the following sections. First, we discuss AI in more detail to articulate the scope of the research within the field and to review a range of prior work related to the problems associated with diminished legibility in this context. Next, we introduce our RtD study including an overview of relevant research into signs and symbols, followed by a reflective account of the design process and its outcomes. Finally, in our discussion and conclusion we reify the paper's contributions; drawing conclusions from the RtD process to shed critical light on prior research and to describe practical implications for the design of systems which utilize or interact with AI.

RELATED WORK

In this section we discuss AI from multiple perspectives in order to map out why legibility for AI is a salient issue for HCI. This includes exploring both how HDI characterizes legibility and the crossover between issues relating to HDI and AI. We review recent HCI literature which relates to AI in order to position the paper's contributions within the field. In addition, we broaden our gaze to consider what lessons HCI may learn from other research communities. Initially, however, we consider how history has created an ambiguity around AI creating a kind of definitional dualism.

The Definitional Dualism of AI

In our introduction we already discussed how AI simultaneously refers to the grand vision of creating a machine with human-level general intelligence *as well as* describing a range of real technologies which are in widespread use today. We might call this AI's definitional dualism. The grand vision describes 'strong' AIs that have a 'general' intelligence—and were such a machine created then the consequences of would be unavoidably monumental [8]. On the other hand, AI also refers to various techniques (e.g. Bayesian networks, deep learning and artificial neural networks) which can routinely be used to create AIs which are 'narrow' or 'weak'. These are employed today to enable a vast array of functionality (e.g. natural language processing, facial recognition, user profiling), which can then be applied in no-end of different

[1] This recent entry in a freely available encyclopaedia provides a comprehensive set of reference material relating to the RtD movement [51].

contexts (e.g. chatbots, surveillance, advertising). Both sets of thinking have shared history, with narrow/weak AI arguably being the result of the community's failed attempts to create strong/general AI. That there is shared history further confuses attempts to distinguish between these two AI pillars, which together manifest as AI's 'definitional dualism'.

Of course, this definitional dualism is not *in itself* problematic, but rather is a symptom of how language and meaning are fluid and change over time. AI is not unique in this sense, for example the similarly buzzword-like term 'Internet of Things' has similar traits. Originally referring to the use of radio frequency identifiers in supply chains it is now simultaneously synonymous with domestic gadgetry, smart cities, and the modern incarnation of Weiser's 'ubicomp' [54]. The linguistic flux of meaning that occurs over time is to be expected, however, in the case of AI the co-evolution of meaning across these two broad interpretations of AI is unbalanced. On the one hand, ethical and moral discussions exploring the potential implications of strong/general AIs are hypothetical, speculative, and rarely profitable—and as such world-leaders are quite open and happy to disclose precisely what the field is achieving. On the other hand, the leaders in weak/narrow AI are corporate entities, where transparency is curtailed by the need to remain competitive. Although the underlying computing techniques used in commercial AI are not secret the detail of how they are incorporated into products and services is rarely made clear. Rather, when AI is publicized in commercial contexts it is often as a marketing or public relations tool [cf. 57]. Hence, AI's dualism which is grounded in the semantics of what the term refers to, actually extends to include the cultures around *how* the term is used too. Although exploring the proposition fully is beyond the scope of this paper, we suggest that corporate utilization of the term AI may be interpreted in terms of fantastical (i.e. strong/general AI) side of the duality which simultaneously obscures or distracts from the extent and ubiquity of how AI is used which, although sometimes banal, has widespread impact on users.

We note that while this paper's internal rhetoric requires that we discuss the distinction between strong/general and weak/narrow AI our interest is principally interested in contemporary, functional and practical uses of AI. Simply put we are concerned with technologies that are in use today, and which are referred to as AI.

Something which all such technologies have in common is their relationship with data. For example, a facial recognition AI such as Amazon Rekognition may be trained on a large, pre-labelled, dataset. By being shown many examples of 'happy' and 'sad' faces, a model learns to tell the difference very efficiently. The resulting AI model may then be incorporated into some other system and process no-end of similar data. In some implementations data that are processed by the AI may be also used as ongoing training data. While each implementation is specific, in all cases AIs process some form of data. The results of this can be remarkable performance given particular tasks (e.g. interpreting brain scans [13], identifying fake smiles [53], devising video game strategies [6]). The data a given AI uses is pivotal to how the AI works and how well it functions. Oftentimes data are created by people. Hence, when considering AI, HDI's interest in how people and data interact with each other is crucial.

Legibility and Human-Data Interaction

HDI is a complementary to HCI, with a focus on individuals, the data that they create, and the algorithms used to analyze them. In addition to the raw data that results from specific interactions HDI takes account of pre-existing or derived data. Analysis of data leads to inferences which result in real-world actions whose effects could be invisible to the individuals to who the data pertains. These inferences may be also collated and then 'fed back' into a system for further analysis. HDI acknowledges the scale of the societal impact stemming from these assemblages of data, analysis and interface and argues to make systems less opaque and to enhance control for individuals. To achieve this HDI is, necessarily, a multidisciplinary endeavor, garnering insights and identifying routes forward aligned with computer science (including HCI), statistics, behavioral economics and sociology. While expansive, the diverse perspectives which inform HDI are distilled into its three core tenets—legibility, agency, and negotiability—and thus the enormity of the HDI challenge is dissected into more manageable chunks.

Legibility refers to what data is collected or processed, how are inferences drawn from it, and what the implications of those inferences are—making a clear distinction between transparency (i.e. not hiding what is going on) and legibility (i.e. making what is going on comprehensible). The agency aspect of HDI is concerned with the capacity for individuals to act within a data system, for example, by being able to transfer one's data, insist it is deleted, or to correct errors in it. Where legibility is concerned with providing information, agency provides the means to do something based on that knowledge. The final attribute—negotiability—explores the broader context in which agency and legibility may manifest, exploring the intricacies of 'societal contracts' around data usage. These are expansive issues including the role of regulation, the need to weigh individual rights against the greater good, and how new cultures around data will change the social context. Whilst HDI's tenets overlap with each other significantly, in this research we focus our interest through the lens of legibility. Straightforwardly, legibility is relevant HCI because the point at which individuals interact with a system is arguably the most obvious useful moment to convey information about the system's implications. Moreover, HCI research around embodied interactions and

data visualization is sympathetic to the aims of HDI's legibility attribute.

Whilst HDI's view of legibility is cast in terms of human relationships with data, as already discussed AI is intrinsically bound up with data too. Hence, the aspiration to explore AI legibility is conceptually very close to HDI legibility. Accordingly, we are interested in exploring questions of what AIs are used to achieve, how they achieve those outcomes, and what the implications of the outcomes might be. Though the role, quality, and type of raw data is a crucial element, *AI* technologies have unique attributes which we explore in the following.

Guidelines for Human-AI Interaction

Reflecting on 20 years of research relating to interactions with AI, Amershi et al. propose and verify principles or guidelines for human-AI interaction. The motivation for the work is to respond to contemporary AI issues such as bias, false-positives, unpredictability, and the impact of whether existence of AIs are visible or "behind the scenes" [2]. Over 150 AI-related design recommendations, collated from prior research, were considered before being distilled into 18 guidelines which aim to be generally applicable. Through an iterative process 49 domain experts validate the guidelines and contextualized them against 20 popular 'AI-infused' systems (the authors use this term—which we adopt in this research—to refer to features which are exposed to the user and harness AI). The 18 guidelines are varied, sometimes referring to the provision of information (e.g. G1, help the user understand what the system is capable of), sometimes covering ethical issues (e.g. G6, ensure undesirable stereotypes and biases are not reinforced), and sometimes describing how a system might be configured (e.g. G17, allow the user to customize what the AI monitors and how it behaves). While the guidelines are meaningfully validated, when viewed alongside the diversity of related work and contextualized with examples from real-world applications, the scope of the such a task is clearly considerable.

Evidenced by the large number (150) of recommendations gleaned, there is a vast range of relevant research to draw upon—for example, insights as diverse as determining the intelligibility of AI models driven by mathematical proof [55] to survey-based-measurement of perceived transparency [46]. Hence, consolidating such variety into *general* guidance is a challenge that is exacerbated by community and disciplinary idiosyncrasies (e.g. Computer Science/Sociology, or HCI/SIGKDD/AAAI). Through the process of validating guidance, these generalization challenges are reflected in the not-insignificant volume of instances where guidelines are not applicable, or they only seem to make sense within specific applications. Moreover, the authors acknowledge that whilst they considered ethics and fairness, the complexity of these concepts far exceeds the straightforwardness of how the guidance is worded (e.g. G6, to mitigate an AI's social bias). Although such well-considered guidelines represent a healthy and significant step towards designing AI systems that are more rigorously human-centered, the designers influence will always be paramount "it is imperative that system designers carefully evaluate the many influences of AI technologies on people and society" [2]. We concur with this sentiment. But notwithstanding a designer's responsibilities, we aim to explore how increased legibility may provide means for users to reap the benefits guidance (e.g. G1, make clear what the system can do) independently.

Transparency, Interpretation and Understandable AI

There is a well-established, and growing, body of work relating to how we might communicate aspects of AI systems to users. An exhaustive review of this landscape is beyond the scope of this paper, but in the following we highlight salient work to articulate the gap in existing research that our own study aims to contribute toward filling.

"If the users do not trust a model they will not use it": focusing on classifiers (e.g. a supervised learning approach using labelled data)—Ribeiro et al. discuss the importance of trusting the predictions that AIs might make [47]. While the assertion that adoption directly correlates to trust is questionable given that oftentimes users do not have a meaningful choice about adopting some technologies [38], there is evidence that acceptance can be enhanced by explaining how a system works [29]. For designers and developers seeking trust "explanations are particularly helpful in identifying what must be done to convert an untrustworthy model into a trustworthy one" [47]. Riberio et al.'s implementation of Local Interpretable Model-agnostic Explanations (LIME) is impressive and demonstrates how it is possible to, in a way which is "locally faithful" (i.e. computable for a specific input), to convey to users how a given classification is rendered. What's more they demonstrate how explaining classifications to users also opens up the opportunity for users to relatively-straightforwardly help make a previously untrustworthy model, trustworthy (where trust equates to its ability to make an accurate classification).

LIME is one example within an extended family of research [e.g. 1,45,55] which seeks to develop the tools and technical foundations that can help explain—for various purposes—what is going on within AI systems' black boxes. Weld and Bansal exploration of "intelligible intelligence" is particularly useful in mapping current efforts to articulate how specific AIs work more comprehensible for designers, developers and users alike. Among the wide array of issues they discuss are: distinguishing the underlying mathematical challenges from the human-focused HCI challenge; unpacking a wide range of reasons why making AI intelligible matters (e.g. legal imperatives, helping humans enhance their own understanding, driving user acceptance, allowing users to control AIs); defining what intelligibility actually refers to;

ranking or quantifying intelligibility; differentiating between intelligible and inherently inscrutable models [55]. While many of these factors are necessary to understand the challenges and drivers of AI legibility, much of the technical detail is likely beyond the grasp of the majority of lay users and therefore is unlikely to be particularly useful as a design element in 'normal' user interfaces. However, Weld and Bansal do raise two highly pertinent points for legibility in their conclusion. First, that a key challenge is the "construction of an explanation vocabulary", acknowledging that, given the scope of AI, this may include relying on some form of generalization. Second, they underscore that explanation is a social process "best thought of as a conversation" [55]. The paper makes it clear that properly grasping how a given AI system works in terms of its technical components is an important factor when attempting to elevate its legibility, but that understanding how one might explain that whilst balancing accuracy, completeness, and accessibility is equally significant.

Focusing more on this communication challenge, in their currently unpublished research for IBM, Arnold et al. consider how documents known as *supplier's declarations of conformity* may be repurposed for AI in the form of *FactSheets* [5]. Building from the assumption that a lack of trust in AI will ultimately stifle its adoption these FactSheets are aimed at developers. The authors rightly point out that we should be mindful that, oftentimes, the way AI is integrated into products is via an API, and hence the developer has no knowledge of how the underlying model works, what data it is trained on, and so on. Moreover, a skills gap tends to exist between those producing and those consuming the AI service, making a critical assessment (by the consumer) even harder. In such circumstances "it becomes more crucial to communicate the attributes of the artifact in a standardized way" [5]. The FactSheets proposed are straightforward and comprise a series of questions under a number of thematic headings which relate to potentially problematic AI issues (e.g., security, explainability, and fairness). The research is presented pragmatically exploring factors which may influence whether such a scheme would be widely adopted. The authors speculate that FactSheets wouldn't need be a legal requirement, but that market forces (e.g. users applying pressure to developers who in turn apply pressure to suppliers) will encourage adoption of FactSheets and ultimately help to avoid an 'AI market for lemons' (e.g. information asymmetry drives uncertainty around quality then the resulting lack of trust harms the market). Although they do not call for blanket regulation, the authors note that there is unlikely to be a one-size fits all solution and that in high-stakes arenas (e.g. health or childcare) FactSheets could become a key part in accreditation. While the FactSheets are described as a 'business to business' tool—as with supplier's declarations of conformity helping AI suppliers and AI developers to develop mutual understanding and trust—their potential role as the foundation in enhancing consumer's understanding of products which have AI incorporated into them is evident. The authors suggest FactSheets may enable initiatives for AI not dissimilar to the *Energy Star* product labeling program, or nutrition labeling on foods.

The success of nutrition labelling in helping to increase consumer awareness about food has helped to popularize the approach when addressing various facets of computing. In one example a 'Dataset Nutrition Label' was developed citing possible benefits including aiding data professionals in selecting the best dataset for their purpose, enhancing quality for those publishing data, and ultimately altering norms toward a more conscientious engagement with data in order to limit possible harms associated with AI [30]. Other similar schemes include schemes to enhance to consumer choice around Internet of Things devices [7]; improving user engagement with privacy policies [32,33]; and to more effectively communicate the efficacy of ranking algorithms [56]. Each of these endeavors struggles with the challenge of providing full and clear information, and the requirement to do so in a succinct and understandable manner. Beyond the challenge of generalization such schemes need to balance the well-evidenced changes in behavior labeling can bring about [14,31] with similarly evidence to suggest the actual changes may not correlate to the intentions [35]. Designing labels for food which are comprehensible and achieve the desired effect is demonstrably difficult. However, the contrastingly easy-to-quantify attributes of food (e.g. saltiness, fattiness, or protein content), when compared to the difficult-to-define attributes of AIs, might suggest that looking to food labeling as a solution to improving AI legibility is a marriage of convenience rather than a rational design.

Rader et al.'s study into algorithmic transparency highlights some of the difficulties in meaningfully researching how one might effectively communicate about AIs [46]. The research uses four functions of transparency—awareness, correctness, interpretability, accountability—and measures the effects of different explanatory interventions on these functions. What becomes clear is that *any* explanation causes users to become more aware of algorithmic influence, but that enhancing users' ability to judge whether an algorithm is behaving correctly, sensibly, or consistently is more of a challenge. However, as noted by the authors, separating generalizable effects from a widespread lack of user base-knowledge (e.g. that most people don't understand the role of algorithms on the Facebook News Feed at all) means that making causal links about awareness interventions is not straightforward. In addition, the findings derived from the walled garden of the Facebook News Feed may be difficult to translate to other scenarios. Nonetheless, the research did seem to suggest there may be a link between explanation of 'what' an algorithm does and user's perception that they can judge when it is performing correctly, and that an explanation of 'how' algorithms work

enhances users' perceived ability to interpret the algorithm's influence. In conclusion Rader et al. raise a thorny issue of user experience; "if the aim is to provide information that users are not aware of, then it seems inherently difficult to ensure that the new information does not violate user expectations" [46]. This final point is, perhaps, indicative of latent societal norms which are waiting to emerge with respect to AI. As the technology becomes domesticated, culture and society may need to adapt to the new reality which AI is helping to create [cf. 39,50].

With this review of related work, we articulate a gap in existing research, that the paper's aim is to contribute toward filling in. The discussion of 'definitional dualism' characterizes the widespread challenge of interpreting what is meant when the term AI is used. Drawing on the fledgling field of HDI we identify the 'legibility' tenet as a key concept sitting at the confluence AI and HCI; in other words, making a user's interaction with data, algorithms, and AI legible should be a concern for HCI research. In our discussion of guidelines for AI and research which attempts to communicate effectively about algorithms, data and AI, what is clear is a widespread and uncontested agreement across research communities that there is a need to be addressed. Evidenced by the of perspectives on these issues it is clear that meaningfully responding to the challenge will take sustained and interdisciplinary effort[2] reflecting a range of research approaches and epistemological perspectives. In the following section we explain how design-led research should play a unique role in this interdisciplinary response to AI's emerging challenges and introduce our RtD study of the space.

DESIGNING FOR LEGIBILITY

While RtD is increasingly used, both in HCI and beyond (the fourth biennial conference for RtD was held in 2019 [cf. 15]), the field is still maturing and "explicit theory about RtD is still in its formative stage" [51]. With this in mind, but not wishing to be swamped by 'pre-paradigmatic' [20] anxieties, we describe our own use of the RtD approach pragmatically and aim to include some core attributes which make a 'good' CHI design paper [cf. 21]. The prototypical designs which we present later in this section represent a 'designerly' [12] response to the challenges that the paper's rhetoric thus far has described. These include the fundamental issue of meaningfully defining AI, how the needs described by HDI must be met by HCI, the (understandable) lack of cohesion across the wide variety of research attempting to understand AI transparency and explanation. The abundance of disparate issues and responses relating to AI's legibility are a good match RtD's attributes; for example, the designerly mindset aiming to unify and integrate multiple perspectives [18] into a coherent conceptual whole, and then embed that unified whole into a tangible output [15,20,51]. By reflecting on this process (e.g. considering the problem space, integrating perspectives on the problem, and producing a tangible response) RtD produces knowledge. It is the rationale for, and details of, that reflexive process are recounted in this section.

Considering AI's Iconography

In addition to RtD's integrative potential, another reason why it is a good fit for this research dilemma is that the problem is—in part—one relating to communication, semiotics, and iconography. These are factors which a design-led inquiry is well-suited to respond to. Early on in this process we searched popular stock imagery and icon databases to consider existing iconography relating to AI. A significant proportion of the imagery returned in our searches (we searched for terms including AI, Machine Learning, Deep Learning) seemed symptomatic of the aforementioned 'definitional dualism' of AI. With the exception of images which offer some visual representation of neural networks (e.g. figure 1a), the vast majority of the imagery offers little to actually communicate how an AI might work. While some iconography might suggest a context that the AI would work in (e.g. figure 1b), no imagery articulated how it might act, or what its implications might be, in that context. The proliferation of robot forms and brain-like structures (e.g. figure 1c, 1d) is representative of the pervasive conflation of contemporary AI techniques/advances in robots and the grand vision of sentient machines. The lack of semantics or communication within the imagery suggests there is scope to develop a visual language which would help to enhance AI legibility; where current offerings evoke AI in the abstract they do very little to help *understanding* about the instance of AI.

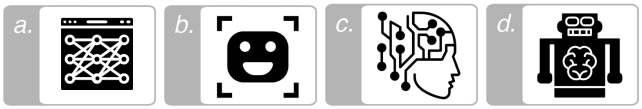

Figure 1. A variety of icons labelled as representing 'AI'.

Icons and other visual cues have been a key tool for aiding interaction with computer systems since the advent of graphical user interfaces [42]. As new modes of interaction have emerged it has fallen to HCI research to help develop icons for them, e.g. auditory icons [19], icons for vehicles [41], and icons for touch-based interactions [4]. Across all these contexts iconography has proven to be a useful tool for encapsulating the complexity of a particular interaction paradigm and symbolizing its attributes for users so that they are aided in knowing how interactions work, and hence can begin to infer implications of the interaction. Visual languages and iconography are aimed *directly* at users, in contrast to much of the research-backed attempts to make

[2] We note that the related work only scratches the surface of the broader issues further relevant reading from a range of other disciplines including Communication [3], Philosophy of Technology [34], Rhetoric [23] and Interdisciplinary Humanities research [9,36].

AI more legible discussed in our related work section. Hence, by developing such a language there is a real opportunity to consolidate suitable elements from the research landscape and package specifically to aid interaction and help communicate how AI is implemented, to users. In sum, given the shortcomings in AI-related iconography, the proliferation of AI into everyday products and services, the consensus that AI legibility should be improved, and the opportunity to repackage a range of existing research in a format digestible by users, there is a clear-cut case for the value of a visual language for AI.

Notes on Signs, Symbols, and Icons

Given the crucial role of signs, symbols and icons in interaction design, HCI scholars have looked to semiotics as a theoretical base for considering icon design. Prior research include taxonomies of icon purpose [40], classification of icon design elements [24], tests to rate intuitiveness of icons [17], studies of icons' ability to maintain a user's attention [37]. The language and theory of semiotics is often used in this context. Of particular note is the Peircean triad. Comprising the *representamen* (the symbol used to represent an idea, e.g. a save icon), the *object* (the actual construct being represented, e.g. data will be saved), and the *interpretant* (the logical implication of the sign, e.g. using this icon will save my data). The relationship between object and representamen can results in three categories of sign: *indexical* signs are those where the signifier (i.e. what is on the sign) is caused by the concept which is being signified (e.g. smoke signifies fire); *symbolic* signs are those which have meaning based solely on convention and may be culturally specific (signs comprised of words are a good example of this, e.g. a 'stop' sign); *iconic* signs have a signifier which resembles the signified (e.g. the paintbrush symbol in a graphics software package) [16]. While these categories are useful as organizing concepts, in reality it "is very rare, and some argue impossible, to find signs that belong solely to one category" [17] and hence apparently straightforward assertions such as "icons are better than indices, and indices are better than symbols" [25] do easily translate to real-world contexts. Considering the past, it is easy to critique failed ideas such as using an icon based on a seashell to represent the 'C-Shell' command processor [24]. Meanwhile the continuing ubiquity of the 'save' icon based on a floppy disk has become an ongoing joke given that floppy disks are antiquated. Hence, concretely theorizing about a given set of symbols or iconography is almost impossible to do. When combined with further difficulty in predicting how or why an icon may become adopted or stay in use, supports the hypothesis that a design-led enquiry may be a useful first step in the process of using icons, symbols or signs to improve AI legibility.

Designing Icons for AI Legibility

In this section we provide a reflexive account of the process of prototyping a number of designs around AI legibility. It is worth noting that various prior research ideas discussed thus far not only make an argument for conducting the RtD study in the first place but also directly inform the designs themselves. The designs included in this section communicate the salient aspects of the process frame the discussion section, but the additional designs which could not be accommodated in the paper itself are provided as supplementary material.

Brains, Brands and Symbols

As a first step in the process we elected to explore the stylistic elements of different visual languages. Within these variations the aim was to keep—in Peircean terms—the object and the interpretant relatively static, while altering the appearance of the representamen based on the following rationale. The three styles are illustrated below.

Figure 2. Contrasting iconographic styles.

In each case the forms were intended to be adapted and augmented to convey additional information, however in the first instance these we designed these icons in order to signify 'AI is present here'. Our first design deliberately utilizes the sort of iconography resulting from AI's definitional dualism (figure 2a) and we referred to as *pictorial*. The second concept explores the use of a brand identity (figure 2b), inspired by the symbology employed by trade organizations such as the *WiFi Alliance*, we referred to this as the *textual* variant. The third approach, which we referred to as the *abstract* variant (figure 2c) and takes cues from the highly recognizable symbolic such as warning signs on roads and laundry labels.

Attributes, Dimensions and Properties

Our next challenge was to consider what the vocabulary of a visual language for AI legibility would or should be. One purpose of the significant review of related work in this paper is to articulate the significant scope of this challenge. Early in the process it was evident that our efforts, while they are well-considered and carefully thought through, would necessarily be but a starting point—seed ideas to be built upon, refined, and adapted and norms and cultures relating to AI become more established. Moreover, this is commensurate with the assertion that RtD-based research tends to be contingent and aspirational [cf. 20].

Given our challenge relates to legibility it was paramount that whatever attributes, dimensions or properties of AI we wished to convey would be comprehensible without a high degree of specialist knowledge. We also recognized the importance of providing space for users to make their own value-judgements based upon their own interpretations; underlining the importance of conveying advisory information rather than qualitative assessment. This is somewhat akin to nutrition labels; for example, while a product's label may inform about high levels of sugar, the

decision about whether to eat it or not ultimately rests with the consumer. Guided by these criteria and based on our review of prior work we arrived six concepts to reflect through the designs (also see figure 3).

- *Presence;* is AI processing taking place by using this service, device or feature? Simply by seeing anyone of our icons (e.g. figure 5) a user can be confident that some form of AI processing is happening, furthermore the placement of the icon (e.g. within a camera app) can be used to suggest to the user which feature within multipurpose devices or software are using AI.
- *Processing Location;* is the AI processing taking place within the device, outside (i.e. in the cloud), or both? While it may not have any direct relevance on the quality of the AI's processing, whether processing is taking place locally or remotely may impact upon user perception of accountability [46].
- *Learning Scope;* is the AI feature static, does it adapt based on usage (e.g. the model is refined based on inputs) or is does it behave as a '2nd order adaptive' system (e.g. is fed by, or feeds into, another AI). Although most AIs function as 'black boxes' involving users in a meaningful 'conversation' [55] which begins with explaining whether the AI will adapt based on using it, and whether that may impact some other 3rd party system, is a straightforward and significant part of that conversation is commensurate with several AI guidelines for articulating how a system may adapt over time [2].
- *Training Provenance;* what is the source of the training data—proprietary, public, or the user themselves? The qualities of the data which help to train an AI are often directly reflected in its behavior, furthermore any ethical issues with data are arguably inherited too, these are demonstrably significant factors for users [5,44,46]. In the future a more granular categorization may be useful, but differentiating between public, private and personal data is a useful first step highlights the value of this information as well as the challenge in obtaining it.
- *Training Data Type;* what data type(s) are used for training this AI, for example visual data, audio data, location data? In a similar vein to the quality of an AI being a function the quality of the data it was trained on, the *type* of data it is trained is a crucial element in reducing the opacity associated with AI black boxes [9].
- *Intrinsic Labor;* through normal use of the AI-enabled feature, is 'work' being done for the AI operator? This factor is somewhat different to the other in that it is hard to ascertain and is inherently subjective. Reflecting on how the monetization of data is driving the commodification of users and their everyday interactions [cf. 43]—this concept strives to communicate the value of users' interactions with an AI, for the AI operator (e.g. Amazon's speech recognition AI may benefit from being used through its Echo devices).

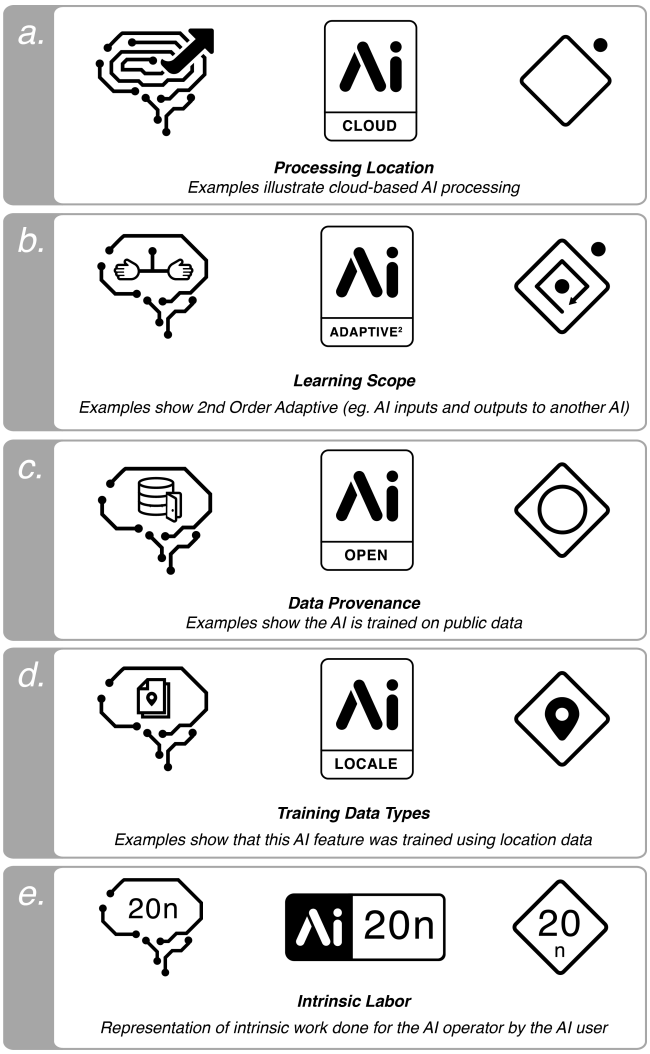

Figure 3. Variants of icons reflecting AI-legibility concepts.

These factors were carefully chosen such that they communicate concepts (i.e. the interpretant) which are objective and thus provide the user with concrete information. Striving for objectivity meant we had to omit designs intended to address some critical aspects of AI; for example, we make no attempt to suggest whether an AI is biased or not. Whilst avoiding these concepts is a shortcoming of our proposals, it is a necessary compromise of the design task at this stage and does not detract from the insights that our study offers. Given our focus is *legibility,* our measure of success is the provision of information to allow users to better form their own opinions (as opposed to imparting value-judgements via constructs such as bias). An exception to this is the concept of 'intrinsic labor' which we included for two reasons. First, the concept resonates with contemporary concerns relating the data it is necessary to provide in order to use many internet services and how those data are often monetized [26]. Defining how to calculate such a value meaningfully is beyond the scope of the paper, but for our purposes we suggest it is a number which explains the amount of monetization stemming from

different AI-infused and allows such systems to be compared to each other numerically. Intrinsic labor could, for example, represent value gained by the AI operator, per hour the AI system is used. The second reason for including intrinsic labor is the study is as a proxy for various hard-to-define but theoretically quantifiable concepts (e.g. fairness or bias) and to demonstrate that whilst such concepts are relatively easily incorporated into a visual language, how useful they are is a factor of how they are calculated and what social meaning they carry.

Reflection, Refinement and Use Cases

Our RtD process hinged around exploring three styles (figure 2) in terms of various attributes (figure 3) in terms of the plethora of related research. This was an iterative process, during which we adapted the designs based on on-going reflection and critique. In supplementary material we provide a fuller account of the design process visually, but in this section, we describe the insights emerging from the iterative reflection and refinement process, ultimately leading toward considering how versions of our visual language might be incorporated into products, devices and services.

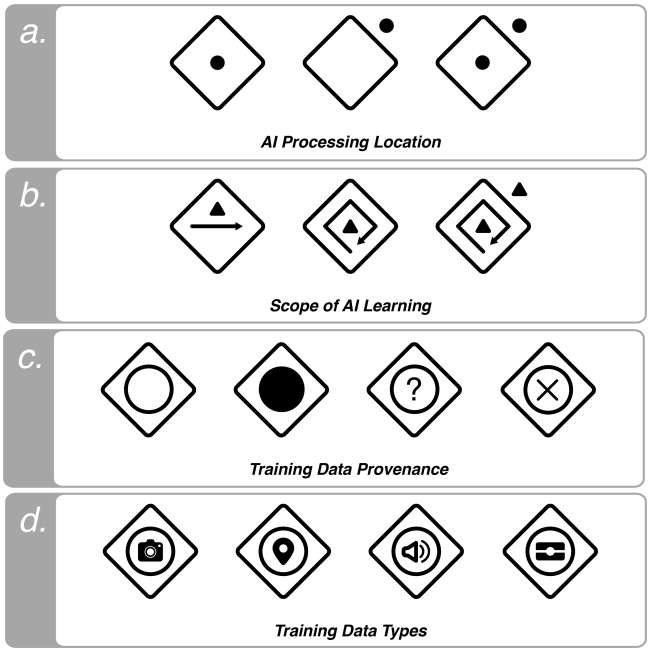

Figure 4. Final iterations of the abstract style into a visual and modular language for AI legibility.

There is an interplay between the visual styles (pictorial, textual and abstract) of our representamen and our relationship with them as designers. Whilst the pictorial AI-brain style (figure 2a) was included in reference to the problematic aspects of AI's definitional dualism, an unexpected benefit of this pictorial visual style was the scope it created to craft iconic (as opposed to indexical or symbolic) imagery (e.g. figure 3c). However, whilst iconic imagery provides an advantage for some concepts for many AI-related constructs iconic imagery difficult to develop because the concepts are so complex or ineffable (e.g. figure 3b). In these cases, it seems sensible to consider the use of symbolic imagery. Another practical issue with the pictorial style is that using multiple signs at the same time becomes problematic. Our more symbolic styles—which we referred to as textual and abstract—address this and are more easily designed in a modular sense (see figure 5).

Our textual design style (figure 2b) was deliberately developed with branding in mind. Trade bodies (e.g. the *WiFi Alliance*) and regulators (e.g. *Conformité Européenne*, or the *Federal Communications Commission*) utilize this kind of branding to make it legible to consumers that products conform to the expected standards (see figure 5a). What the actual standards which organizations using these marks adhere to differs wildly across domains, but the brand marks play a key role in communicating to users, a guarantee of compatibility and conformity to minimum safety considerations. While such approaches benefit from the so-called "feature positive effect" [31] (i.e. the presence of any kind of label encourages a more critical though process) it is difficult to combine this with *reliably* conveying the information which is intended to enhance legibility [35]. Moreover, that well-supported schemes of this type have failed to be widely adopted in the technology space [cf. 7] suggests that a reconfiguration of industry, regulation, and consumer-demand is needed for such an approach to succeed. Insofar as communicating the actual attributes and AI concepts, this approach was symbolic, but focused solely on text. Individual icons employ keywords situated alongside the brand in order to communicate key concepts (e.g. 'Adaptive^{2}' to represent 2nd order adaptive). Whilst we envisaged some scope to modularize this design (e.g. using multiple keywords in a single instance) there is a limit to how much textual information is viable to present in a single instance (the example in figure 5a offers a compromise).

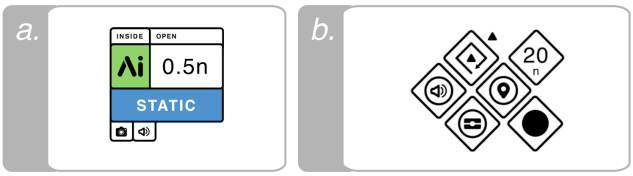

Figure 5. Modular examples of multiple icons used together.

The abstract design style (figure 2c) became the main focus of our iterations as it offered the most scope to engage with the issues and draw out insights in a designerly manner. This set of designs also seemed to minimize the problematic aspects of the other approaches. For example, although iconic imagery could be incorporated into the abstract style if appropriate, it did not demand pictorials necessarily. In addition, the textual designs derived much of their value by association with a brand-like identity, and the strength of that identity could potentially move the focal point from legibility and toward identifiability (of the brand). Finally, the form of the abstract icons lends itself to

modularity, which in turn opens up the potential for the individual designs to be used together to produce a much richer meaning—as vocabulary and grammar come together to form a language.

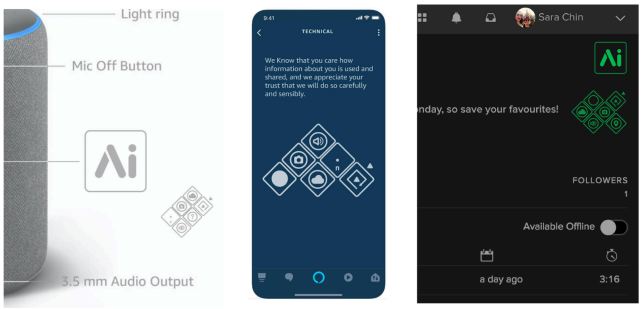

Figure 6. Example use cases.

A number of visual concepts were explored through the iterations (please refer to the supplementary material to review these). In the most refined version abstract icons form a language where small circles denote AI processing (figure 4a), triangles denote AI learning (figure 4b), symbols inside of large circles represent the various types of and provenance of training data (figures 4c and 4d). Used together and in context these abstract icons highly how specific features, services and products interactions with AI can be made more legible (figure 6). In line with the expectations one might expect for an RtD exploration the insights emerging from the work are contingent on the ongoing AI innovation, yet they do proffer a unique perspective on the complex and nuanced problem space of legibility for AI—through the study we aim to surface and make tangible some aspects of a multi-dimensional challenge.

DISCUSSION

The design proposals in this paper are, by no means, intended to solve or conclude the challenge of making AI legible. Rather this work hope to make a coherent argument for exploring specific aspects of the challenges in a unique, and design-led, manner. With a set of interlinked challenges as expansive of those associated with AI it is paramount that research deals with the technical and theoretical detail *but also* with the pragmatic and practical issues, which ultimately will have a significant impact on usability and acceptability of AI-infused products. The thorough review of related AI research in this paper aims to highlight this gap, and thus articulates the motivation for this research. Meanwhile, the RtD approach allows us to utilize the designs presented in this paper to uniquely reflects multiple facets of prior research, and demonstrate them straightforwardly, by focusing around the actual point of interaction between users and AIs (as opposed to exploring tools to help developers and designers [e.g. 2,5]). By making AIs legible to users their interactions with AI need not be 'inert'. In this sense design-led research such as is presented in this paper seems apt to act as a filter to sense-check how state-of-the-art research into HCI and AI may be applied to everyday interactions and circumstances. Based on our exploration it seems that striving for an accessible visual communication approach forces a 'translation' from, for example, fundamental work in what constitutes intelligibility [55] into an accessible constructs such as those represented in our designs. RtD has proved an excellent means to understand, and begin to explore responses to this issue, but further work is certainly necessary.

We believe this study will provide a firm and worthy foundation for future empirical research. Using these designs within the context of Human-Centered Design processes for specific AI applications would help to triangulate our insights with user perspectives, interrogating what factors relating to AI are important to users and what types of visual language are most understandable. Similarly, it would be fruitful to explore, with AI domain experts, the breadth of opinions regarding what information may be useful to represent visually in order to enhance AI legibility (and conversely what approaches to be wary of and potentially detrimental). Finally, given that the impact and ubiquity of AI is largely driven by the commercial entities we see a significant value in understanding the market forces which may underpin adoption of a scheme for enhanced AI legibility by working with the purveyors of AI services as well as policy makers. Whilst all of these more traditional avenues for research will provide incremental steps, with empirical support, toward more legible AI systems, we suggest that design-led work such as presented in this paper should play a key role in bridging disparate disciplinary perspectives and providing a tangible means to communicate and explore concerns of users. While design-led research, as per the RtD school of thought, is always epistemologically distinct to 'scientistic' positivism—the 'aspirational and contingent' [20] knowledge and insights such processes result in should become an established part of the AI research milieu. Notwithstanding RtD's occasionally challenging epistemology, the unique contribution that design-led inquiries can play for HCI a key step toward understanding and responding to the emerging reality of living with AI and making our relationships with them more legible.

REFERENCES

[1] Saleema Amershi, Max Chickering, Steven M. Drucker, Bongshin Lee, Patrice Simard, and Jina Suh. 2015. ModelTracker. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 337–346. DOI:https://doi.org/10.1145/2702123.2702509

[2] Saleema Amershi, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, Eric Horvitz, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, and Paul N. Bennett. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing*

Systems - CHI '19, 1–13. DOI:https://doi.org/10.1145/3290605.3300233

[3] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20, 3 (March 2018), 973–989. DOI:https://doi.org/10.1177/1461444816676645

[4] Timo Arnall. 2006. A graphic language for touch-based interactions. In *Mobile Interaction with the Real World*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.9464&rep=rep1&type=pdf#page=18

[5] Matthew Arnold, Rachel K. E. Bellamy, Michael Hind, Stephanie Houde, Sameep Mehta, Aleksandra Mojsilovic, Ravi Nair, Karthikeyan Natesan Ramamurthy, Darrell Reimer, Alexandra Olteanu, David Piorkowski, Jason Tsay, and Kush R. Varshney. 2018. FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity. (2018). Retrieved from http://arxiv.org/abs/1808.07261

[6] Kai Arulkumaran, Antoine Cully, and Julian Togelius. 2019. AlphaStar: An Evolutionary Computation Perspective. (February 2019). DOI:https://doi.org/10.1145/3319619.3321894

[7] John Blythe and Shane Johnson. 2018. *Rapid evidence assessment on labelling schemes and implications for consumer IoT security*. Retrieved from https://www.gov.uk/government/publications/rapid-evidence-assessment-on-labelling-schemes-for-iot-security

[8] Nick Bostrom. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

[9] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3, 1 (January 2016), 205395171562251. DOI:https://doi.org/10.1177/2053951715622512

[10] Stephen Cave and Seán S. ÓhÉigeartaigh. 2018. An AI Race for Strategic Advantage. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society - AIES '18*, 36–40. DOI:https://doi.org/10.1145/3278721.3278780

[11] Andy Crabtree and Richard Mortier. 2015. Human Data Interaction: Historical Lessons from Social Studies and CSCW. In *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway*. Springer International Publishing, Cham, 3–21. DOI:https://doi.org/10.1007/978-3-319-20499-4_1

[12] Nigel Cross. 2011. *Design Thinking: Understanding How Designers Think And Work*. Bloomsbury.

[13] Yiming Ding, Jae Ho Sohn, Michael G. Kawczynski, Hari Trivedi, Roy Harnish, Nathaniel W. Jenkins, Dmytro Lituiev, Timothy P. Copeland, Mariam S. Aboian, Carina Mari Aparici, Spencer C. Behr, Robert R. Flavell, Shih-Ying Huang, Kelly A. Zalocusky, Lorenzo Nardo, Youngho Seo, Randall A. Hawkins, Miguel Hernandez Pampaloni, Dexter Hadley, and Benjamin L. Franc. 2019. A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18 F-FDG PET of the Brain. *Radiology* 290, 2 (February 2019), 456–464. DOI:https://doi.org/10.1148/radiol.2018180958

[14] Andreas Drichoutis, Panagiotis Lazaridis, and Rodolfo Nayga Jr. 2006. Consumers' use of nutritional labels: a review of research studies and issues. *Academy of marketing science review* (2006).

[15] Abigail C. Durrant, John Vines, Jayne Wallace, and Joyce S. R. Yee. 2017. Research Through Design: Twenty-First Century Makers and Materialities. *Design Issues* 33, 3 (July 2017), 3–10. DOI:https://doi.org/10.1162/DESI_a_00447

[16] Jennifer Ferreira, Pippin Barr, and James Noble. 2002. The Semiotics of User Interface Redesign. In *Proceedings of the Sixth Australasian conference on User interface*, 47–53.

[17] Jennifer Ferreira, James Noble, and Robert Biddle. 2006. A case for iconic icons. In *Conferences in Research and Practice in Information Technology Series*, 87–90.

[18] Lois Frankel and Martin Racine. 2010. The Complex Field of Research: for Design, through Design, and about Design. *Design Research Society* (2010).

[19] William Gaver. 1986. Auditory Icons: Using Sound in Computer Interfaces. *Human-Computer Interaction* 2, 2 (June 1986), 167–177. DOI:https://doi.org/10.1207/s15327051hci0202_3

[20] William Gaver. 2012. What should we expect from research through design? In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, 937–946. DOI:https://doi.org/10.1145/2207676.2208538

[21] William Gaver and Kristina Höök. 2017. What makes a good CHI design paper? *Interactions* 24, 3 (2017), 20–21. DOI:https://doi.org/10.1145/3076255

[22] Elizabeth Gibney. 2016. Google AI algorithm masters ancient game of Go. *Nature* 529, 7587 (January 2016), 445–446.

DOI:https://doi.org/10.1038/529445a

[23] Karamjit S. Gill. 2016. Artificial super intelligence: beyond rhetoric. *AI & Society* 31, 2 (May 2016), 137–143. DOI:https://doi.org/10.1007/s00146-016-0651-x

[24] David Gittins. 1986. Icon-based human-computer interaction. *International Journal of Man-Machine Studies* 24, 6 (June 1986), 519–543. DOI:https://doi.org/10.1016/S0020-7373(86)80007-4

[25] Joseph Gogeun. 1993. On Notation. In *Tools10: Technology of Object-Oriented Languages and Systems*, 47–53.

[26] Samuel Greengard. 2018. Weighing the impact of GDPR. *Communications of the ACM* 61, 11 (October 2018), 16–18. DOI:https://doi.org/10.1145/3276744

[27] Hamed Haddadi, Richard Mortier, Derek McAuley, and Jon Crowcroft. 2012. *Human Data-Interaction*.

[28] P Hayes and K Ford. 1995. Turing test considered harmful. *International Joint Conference on Artificial Intelligence* (1995), 972–977. Retrieved from http://www.csee.umbc.edu/courses/471/papers/hayes95.pdf

[29] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work - CSCW '00*, 241–250. DOI:https://doi.org/10.1145/358916.358995

[30] Sarah Holland, Ahmed Hosny, Sarah Newman, Joshua Joseph, and Kasia Chmielinski. 2018. The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards. May (2018). Retrieved from http://arxiv.org/abs/1805.03677

[31] Frank R. Kardes. 1988. Spontaneous Inference Processes in Advertising: The Effects of Conclusion Omission and Involvement on Persuasion. *Journal of Consumer Research* 15, 2 (September 1988), 225. DOI:https://doi.org/10.1086/209159

[32] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security - SOUPS '09*, 1. DOI:https://doi.org/10.1145/1572532.1572538

[33] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. 2010. Standardizing privacy notices. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, 1573. DOI:https://doi.org/10.1145/1753326.1753561

[34] Asle H. Kiran and Peter-Paul Verbeek. 2010. Trusting Our Selves to Technology. *Knowledge, Technology & Policy* 23, 3–4 (December 2010), 409–427. DOI:https://doi.org/10.1007/s12130-010-9123-7

[35] Joerg Koenigstorfer and Hans Baumgartner. 2016. The Effect of Fitness Branding on Restrained Eaters' Food Consumption and Postconsumption Physical Activity. *Journal of Marketing Research* 53, 1 (February 2016), 124–138. DOI:https://doi.org/10.1509/jmr.12.0429

[36] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (June 2018), 205395171875668. DOI:https://doi.org/10.1177/2053951718756684

[37] Hsuan Lin, Yu-Chen Hsieh, and Fong-Gong Wu. 2016. A study on the relationships between different presentation modes of graphical icons and users' attention. *Computers in Human Behavior* 63, (October 2016), 218–228. DOI:https://doi.org/10.1016/j.chb.2016.05.008

[38] Joseph Lindley, Sara Canizzaro, Rob Procter, and Paul Coulton. 2019. Adoption and Acceptabillity. In *Cybersecurity of the Internet of Things*, K Pothong, I Brass and M Carr (eds.). PETRAS IoT Research Hub. Retrieved from https://www.petrashub.org/petras-stream-report/

[39] Joseph Lindley, Paul Coulton, and Miriam Sturdee. 2017. Implications for Adoption. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 265–277. DOI:https://doi.org/10.1145/3025453.3025742

[40] Xiaoyue Ma, Nada Matta, Jean-Pierre Cahier, Chunxiu Qin, and Yanjie Cheng. 2015. From action icon to knowledge icon: Objective-oriented icon taxonomy in computer science. *Displays* 39, (October 2015), 68–79. DOI:https://doi.org/10.1016/j.displa.2015.08.006

[41] Aaron Marcus. 2002. Information Visualization for Advanced Vehicle Displays. *Information Visualization* 1, 2 (June 2002), 95–102. DOI:https://doi.org/10.1057/palgrave.ivs.9500016

[42] Aaron Marcus. 2003. Icons, symbols, and signs. *interactions* 10, 3 (May 2003), 37. DOI:https://doi.org/10.1145/769759.769774

[43] Evgeny Morozov. 2013. *To Save Everything Click Here: Technology, Solutionism and the Urge to Fix Problems That Don't Exist*. Allen Lane Penguin Books.

[44] Richard Mortier, Hamed Haddadi, Tristan Henderson, Derek McAuley, and Jon Crowcroft.

2014. Human-Data Interaction: The Human Face of the Data-Driven Society. *SSRN Electronic Journal* (2014). DOI:https://doi.org/10.2139/ssrn.2508051

[45] Kayur Patel, Naomi Bancroft, Steven M. Drucker, James Fogarty, Andrew J. Ko, and James Landay. 2010. Gestalt. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10*. DOI:https://doi.org/10.1145/1866029.1866038

[46] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–13. DOI:https://doi.org/10.1145/3173574.3173677

[47] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 1135–1144. DOI:https://doi.org/10.1145/2939672.2939778

[48] Neelima Sailaja, Andy Crabtree, and Phil Stenton. 2017. Challenges of using Personal Data to Drive Personalised Electronic Programme Guides. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 5226–5231. DOI:https://doi.org/10.1145/3025453.3025986

[49] Thomas Albert Sebeok. 2001. *Signs: An Introduction to Semiotics*. University of Toronto Press.

[50] Roger Silverstone. 2006. Domesticating domestication. Reflecting on the life of a concept. In *Domestication Of Media And Technology*, Thomas Berker, Maren Hartmann, Yves Punie and Katie Ward (eds.). Open University Press, 229–247.

[51] Pieter Stappers and Elisa Giaccardi. 2019. 43. Research through Design. In *The Encyclopedia of Human-Computer Interaction, 2nd Ed.* Retrieved September 6, 2019 from https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/research-through-design

[52] A. M. TURING. 1950. I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind* LIX, 236 (1950), 433–460. DOI:https://doi.org/10.1093/mind/LIX.236.433

[53] Hassan Ugail and Ahmad Al-dahoud. 2019. A genuine smile is indeed in the eyes – The computer aided non-invasive analysis of the exact weight distribution of human smiles across the face. *Advanced Engineering Informatics* 42, February (October 2019), 100967. DOI:https://doi.org/10.1016/j.aei.2019.100967

[54] Mark Weiser. 1999. The computer for the 21 st century. *ACM SIGMOBILE Mobile Computing and Communications Review* 3, 3 (July 1999), 3–11. DOI:https://doi.org/10.1145/329124.329126

[55] Daniel S. Weld and Gagan Bansal. 2019. The challenge of crafting intelligible intelligence. *Communications of the ACM* 62, 6 (May 2019), 70–79. DOI:https://doi.org/10.1145/3282486

[56] Ke Yang, Julia Stoyanovich, Abolfazl Asudeh, Bill Howe, HV Jagadish, and Gerome Miklau. 2018. A Nutritional Label for Rankings. In *Proceedings of the 2018 International Conference on Management of Data - SIGMOD '18*, 1773–1776. DOI:https://doi.org/10.1145/3183713.3193568

[57] CES Day One: AI Is Everywhere - SyncedReview - Medium. Retrieved September 6, 2019 from https://medium.com/syncedreview/ces-day-one-ai-is-everywhere-6b13f3999596