DRS2020
BRISBANE, 11–14 AUG
SYNERGY

Signs of the Time: Making AI Legible

Joseph Lindley*, Haider Ali Akmal, Franziska Louise Pilling, Paul Coulton[a]

[a] ImaginationLancaster, Lancaster University, United Kingdom
*Corresponding author e-mail: j.lindley@lancaster.ac.uk

doi: https://doi.org/10.21606/drs.2020.237

Abstract: Artificial Intelligence (AI) is becoming widespread. There are many benefits associated with AI, but it's adoption brings challenges relating to fairness, bias, and transparency. Such issues are particularly hard to address because conventions that highlight when an AI is present, how it works, and the consequences of using are not yet established: AI has a legibility problem. Design-led research can play a key role in exploring this challenge. Applying Research through Design (RtD) this paper explores AI legibility in three ways: (1) explaining why it makes sense to address AI legibility with design; (2) the presentation of prototypical icons designed to enhance AI legibility; (3) experimenting with how the icons may be used in the context of signage relating to potential applications of AI. Via these three lenses the paper argues that design's role in improving AI legibility is critical.

Keywords: Artificial Intelligence, Icon Design, Research through Design.

1. Introduction

Humanity's fascination with artificial life is long-lived, appearing in ancient mythology (e.g. *Galatea*, *Talos*) and more modern fiction alike (e.g. *The Creature* in Shelley's *Frankenstein*, *HAL9000* in Arthur C. Clarke's *2001*). Arguably, however, it was Alan Turing's seminal research question "Can machines think?" (Turing, 1950) that gave rise to the field and the technologies that we now call AI. In the 70 years since Turing posed the question, AI has experienced cycles of inflated expectation and troughs of disillusionment. While the ethical and technical complexities of Artificial *General* Intelligence (AGI) remain as ontologically challenging as ever, the AI field has developed an array of powerful computing techniques including Neural Networks, Expert Systems and Machine Learning. Facilitated by a growing abundance of data, cheap computing power, and advanced data science, these techniques—in particular Machine Learning—have become come widespread. Whilst these AIs excel at pattern recognition and prediction tasks, we have not created any 'thinking' machines, however, there are plenty of reasons why we should put time and effort into thinking about *them*.

The AGIs that appear in fiction love, live, and fight. These emotive characteristics belie how mundane most applications of AI actually are. AI has become key to national strategies (cf. Hall & Pesenti, 2017; *Industrial Strategy*, 2017). This is *not* because super intelligent AGIs are likely to emerge in the near future, but rather that the significant disruptive economic potential of AI has begun to be realised. In light of this we might cast AIs as innovation engines which are fuelled by data, and given this fuel is abundant and cheap it is no surprise the engines are running apace. Notwithstanding the prevailing rhetoric that AI is a proximate future, or 'just around the corner' (Lindley, Coulton, & Sturdee, 2017), applications of AI are already ubiquitous. AI features are integral to activities such as shopping, dating, banking—even the simple act of typing using a predictive keyboard (as in a smartphone).

These ingredients combine to make a cocktail of problematic aspects relating to AI. First, the disconnection between AI's popular vision of intelligent robots, and the reality of faceless and non-cognisant algorithms, is rhetorically dissonant. This reduces the legibility of devices and services which use AI (Gill, 2016). Second, AIs reflect the data which they are trained on, and those datasets are often unrepresentative, inaccurate or biased (Amershi et al., 2015, 2019)—qualities reflected in the AI's trained on them. Third, despite a variety of efforts to make AIs explainable, for most of their users they remain 'black boxes' (Ananny & Crawford, 2018; Ribeiro, Singh, & Guestrin, 2016). Nonetheless AI is employed all around us; AI, and the data which define them, are palpably altering the world in which we live (Lanier, 2013; cf. Morozov, 2013).

In this work we adopt Research through Design (Frayling, 1993; Gaver, 2012)—forthwith 'RtD'—as a means to explore the challenge of AI's legibility. RtD is an apt framework to AI legibility with for several reasons. Prior research and efforts to make AI legible are interdisciplinary[1]. Meanwhile the Design's inherently integrative and generative character (Cooper et al., 2018) provides an opportunity to combine salient aspects of different disciplinary approaches through practice. Issues arising from the adoption of AI form a multifaceted challenge, but it is one that Design-led research is uniquely equipped to deal with by meaningfully combining the theoretical and technical attributes of AI with human-centred and social concerns, and transmuting these varied perspectives into accessible, tangible and novel insights. Integrating the design and research processes through RtD provides a unique opportunity to reify AI's challenges, testing, triangulating, and integrating disciplinary-diverse perspectives. With this in mind it is important to consider that the designs presented in this paper are intended for a research audience as an instrument of RtD, as opposed to design proposals that we intend to be immediately ready for a lay audience or to be adopted in the wild. Operating in this context, RtD processes apply specific design constraints to what Buchannan refers to the "fundamental indeterminacy in all but the most trivial design problems" (Buchanan, 1992). Exploring expansive problem spaces, allowing conception of problem and solution to co-evolve, RtD is uniquely well equipped to

[1] For example, notable research endeavours in Human-Computer Interaction (Amershi et al., 2019), Communication (Ananny & Crawford, 2018), Philosophy of Technology (Kiran & Verbeek, 2010), Rhetoric (Gill, 2016) and Interdisciplinary Humanities research (Burrell, 2016; Lee, 2018).

make sense out of the overwhelming scope of the problem space. In the case of this research the part of the problem space we explore is the *legibility* of AI systems and our constraint is to do so by developing a visual language for enhancing AI legibility. We elaborate on both the concept of legibility and the rationale for developing a visual language in the subsequent sections.

The paper proceeds as follows. In section 2, we have an introduction to different forms of contemporary AI, a brief overview of interdisciplinary AI research programmes, and some notes on what we mean by AI legibility. In section 3 we discuss existing iconography relating to AI and situate our project in relation to semiotics. Section 4 introduces our designs, explaining the design process for the icons and subsequently deploying them in the context of public information signage. In section 5 we conclude with a discussion of the RtD process, highlighting contingent findings, limitations, and future research.

2. What is AI anyway and why should it be legible?

In order to acclimatise readers who are unfamiliar with AI in this section we situate the paper with a brief history and discussion of the state of contemporary AI research. By supposing that the process by which humans learn involves logically processing available data—something which computers do with aplomb—Turing's seminal work became the grand challenge of AI. To achieve the challenge, we'd just require the relevant data, and knowledge of how the learning algorithm works. The famous Turing Test (or 'Imitation Game') was posited as a means to test whether or not AI had been achieved. The game goes thus: in a conversation if a human cannot determine whether they are talking to another human or a computer then we have achieved the grand challenge of AI and proven that, yes, computers can 'think'. Aspiring to pass this test has been a key driver of AI innovation, however it is also somewhat problematic, it suffers as a result of what we term AI's 'definitional dualism'. On one side of this dualism we note the ubiquitous use of AI techniques in narrow use cases—these machines do not think. Meanwhile on the other is the fact that these the techniques were developed in an attempt to create AGI—machines which *do* think. Although this has been productive, resulting in a many very effective computing techniques which we call AI, at the same time the historic connotations of the term AI evokes un-realistic perceptions and mean that AI is judged by unrealistic criteria (Hayes & Ford, 1995), ultimately resulting in a dangerous rhetorical dissonance (Cave & ÓhÉigeartaigh, 2018).

Confounding the definitional dualism of AI, further factors make a clear delineation of what we mean by AI even harder. The recent proliferation of smart speakers and voice agents (e.g. *Siri*, *Alexa*, and Google's *Assistant*) which are routinely referred to as 'AI', thus evoking the image of an intelligence in one's house (when the reality is that they are fairly rudimentary devices utilising AI for voice recognition) highlights this. Elsewhere Apple's 'bionic' chip is described as dedicated AI hardware, and Huawei highlight AI as a unique selling point in their recent handsets, yet in each case, what is really going is a conflation of the two sides of the dualist equation. Notwithstanding the hyperbole of AI's dualism, for the vast majority of the paper we will be dealing with the mundane and non-fantastical form of AI which ubiquitously exists today.

With AI's dualism codified and the field's history acknowledged, we have still not proffered a simple explanation for what we mean by AI. This is because such a reductive account doesn't make much sense; there are in fact many interrelated techniques, use cases, and applications which are referred to as AI. While reviewing all of these, and their relationships is beyond the scope of this work, here we aim to give a pragmatic sense of the space by considering its history.

In the 1950s much effort was put into 'symbolic reasoning' an approach which encodes a hypothesis into logic, generating a tree which can then be searched algorithmically. This was thought of as a model for human reasoning. The approach, perhaps inspired by the Turing test, was applied to understanding and synthesising natural language. Later, in the 1970s AI met the physical world and researchers began to try and make robots which utilised AI. However, by the 1980s optimism around AI had subsided; while the various techniques were viable there was not enough available data, storage, or computing power to make them work properly. Around the same time 'expert systems' became popular. Rather than building search trees based on logics and hypotheses, these systems encoded human knowledge into much smaller decision trees. This negated the issues associated with storage and computing power, and in certain domains these systems were hugely successful. By the late 1990s and early 2000s, as predicted by Gordon Moore, computing power had doubled roughly every 2 years and modern computers were fast enough to properly run AI software. At the same time storage became much cheaper, connectivity (e.g. broadband, WiFi, 3G, 4G, etc) faster and more available, many everyday services were digitised, and the Internet of Things (IoT) became a reality. These factors together have precipitated a rapid and widespread adoption of AI. While a wide range of techniques and methods make up modern AI, perhaps the most significant is Machine Learning (ML). The family of techniques which make up ML allow systems to perform particular tasks (e.g. learning to recognise cats) by learning from patterns in data and ML has been so significant for AI that the two terms are now often used interchangeably.

It is the vagaries of ML which have given rise a raft of contemporary AI concerns relating to understanding bias (Rader, Cotter, & Cho, 2018), fairness (Cave & ÓhÉigeartaigh, 2018; Lindley et al., 2019), and transparency (Ananny & Crawford, 2018; Weld & Bansal, 2019). The gravity of these issues is such that they are attracting multidisciplinary research effort. For example, Human-Computer Interaction scholars are striving to develop guidelines for designing AI systems (Amershi et al., 2019), computer scientists are developing technical methods to provably quantify bias (Ribeiro et al., 2016), and emerging design theories such as 'More-Than-Human Centred Design' update our dogmas for a world where technology is entwined in with society (P. Coulton & Lindley, 2019) as well as a number of other dalliances between contemporary Design Research and Philosophy (Lindley, Akmal, & Coulton, 2020; e.g. Redström & Wiltse, 2019)[2].

The interdisciplinary 'Human-Data Interaction' (HDI) field (Haddadi, Mortier, McAuley, & Crowcroft, 2012) is particularly salient for this work as it frames the concept of *legibility*. HDI proposes three tenets for understanding our relationships with data (and by extension our relationship with AI); *agency, negotiability, and legibility*. In HDI terms legibility is quite distinct from transparency and instead refers a user's ability to comprehend how a system works. The agency aspect of HDI is concerned with the capacity for individuals to act, for example being able to decide not to participate based on comprehension. The final attribute—negotiability—explores the broader context in which agency and legibility may manifest, exploring the intricacies of 'societal contracts' relating to data or AI systems. These are expansive issues, but also have significant overlaps, however, in this paper we describe an RtD project which explores the challenge of AI legibility. In addition to combining aspects of the aforementioned AI research, the work also responds to the current lack of legibility for iconography that is currently associated with AI.

3. Iconography and AI

From religious imagery to calligraphy, iconography is a broad term, in this paper we use it to describe the small graphics used in computing to represent programs, features, or options (cf. Ferreira, Barr, & Noble, 2002)—*icons*. In order to understand how AI (and related concepts such as ML and Neural Networks) tend to be represented we searched a variety of image repositories for icons representing AI. It was evident that definitional dualism is echoed in the image libraries; there is a wide range of brain-like structures, robots, and a proliferation of imagery which evokes conscious, feeling, or thinking machines (see figure 1a, 1c, 1d). There was also a lack of imagery which explains how AI works, or what context it is working in—with two notable exceptions. Neural Networks are commonly depicted as layers or networks of nodes (see figure 1a), to an educated reader this may indicate something about how a particular AI works, however even this gives rise to new questions (e.g. how many layers does the network have, what data are processed, is it an adaptive network?). Similarly, some icons, provide the reader with information relating to the domain of use. In the case of figure 1b we can easily determine that the AI is used to enable facial recognition

[2] We note, as one of our reviewers rightly points out, this is a fast-moving field—we would encourage readers to search for up to date information, both within and outwith the academic realm.

(however, we have no idea what sort of AI-enabled machine vision system is in use).

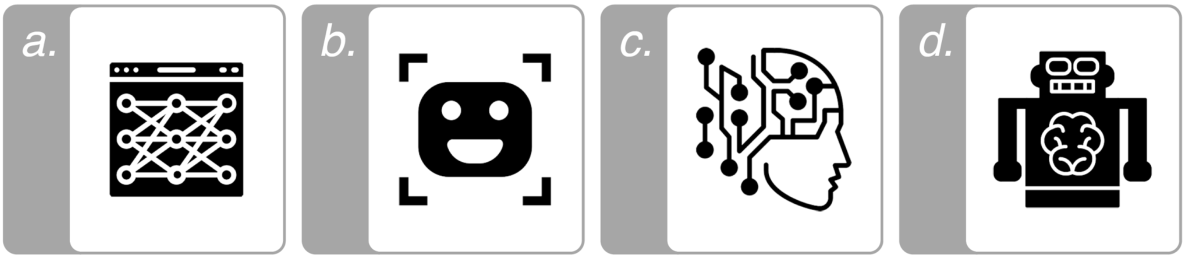


Figure 1. Examples of AI iconography.

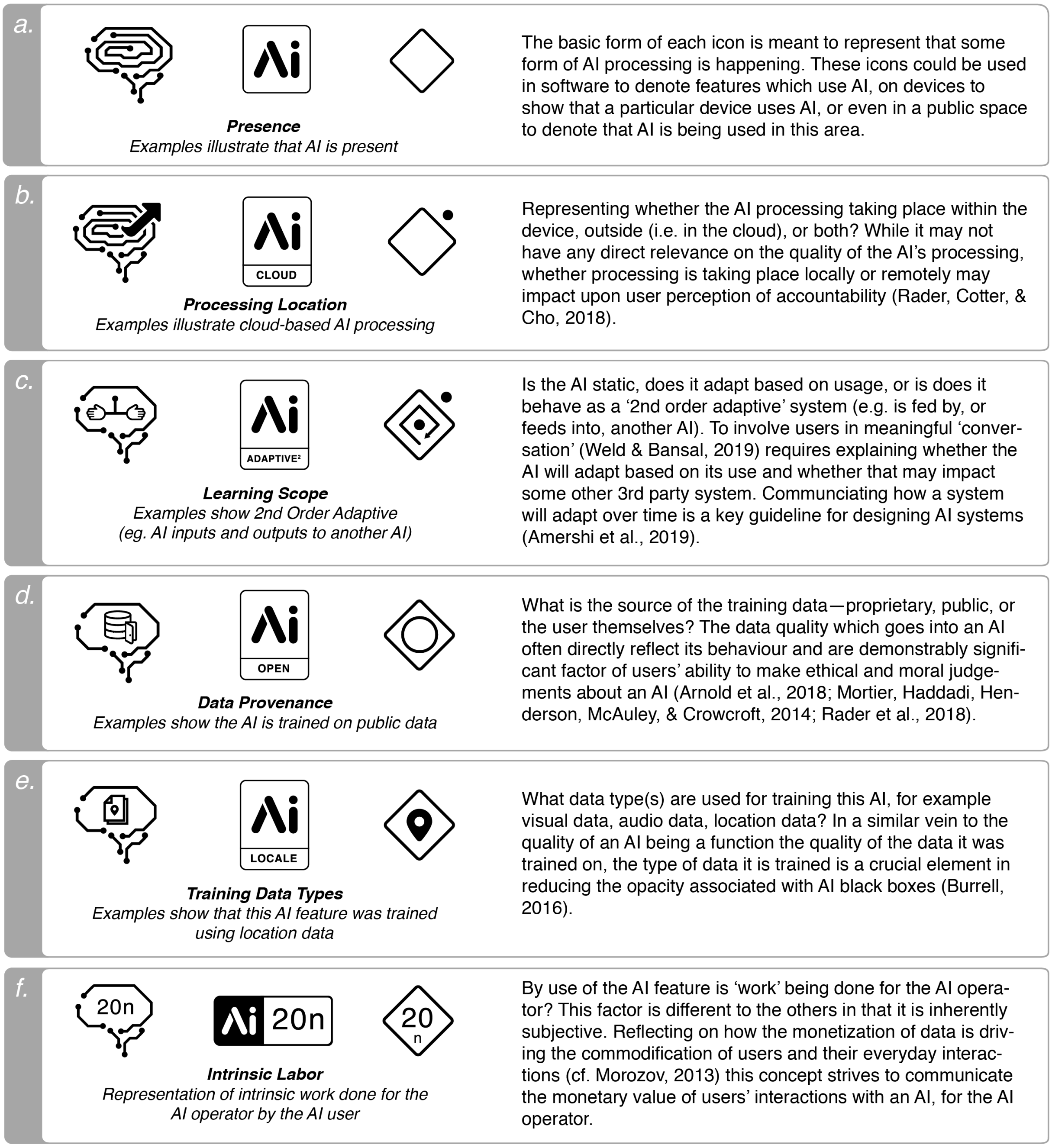
Research into icons is diverse; classifying the icon purpose (Ma, Matta, Cahier, Qin, & Cheng, 2015), deconstruction of icon elements (Gittins, 1986), evaluation metrics for intuitiveness (Ferreira, Noble, & Biddle, 2006). Frequently research intersects with semiotic theory such as the Peircean triad. The triad comprises the *representamen* (a symbol used to represent an idea, e.g. a 'folder' icon); the *object* (the actual construct being represented, e.g. individual 'files' organised into a 'folder'); and the *interpretant* (the sign's implication, e.g. all files in the folder can be moved around together). These constructs are used together to deconstruct different categories of sign (Ferreira et al., 2002). *Indexical* signs are those where the 'signifier' is the result of the concept appearing on the sign (e.g. smoke signifies fire); *symbolic* only have meaning by convention (e.g. a 'stop' sign); when the signifier looks like the signified it is described as an *iconic* sign (e.g. paintbrush tool in graphics software). While this semiotic view of icons is a handy conceptual lens, and provides us with a language to describe the icons with, in reality "is very rare, and some argue impossible, to find signs that belong solely to one category" (Ferreira et al., 2006).

The majority of AI icons have representamen which tie into the fantastical 'killer robot' side of AI's dualism—brains, robots, etc. Hence the interpretant is misleading. On the occasions when interpretant-clarity is increased the sense of the object tends to be sacrificed (e.g. figure 1b is clearly about facial recognition, but with no sense of *how* or *why* that system works). Whilst category-mixing is normal in the Peircean view of signs, in the case of AI, the combination of category mixing and lacking conventions or cultural understanding mean that the majority of AI iconography at best indicative and at a worst misleading. These shortcomings in the current state of AI's iconography highlight the space that this RtD exploration seeks to occupy; a visual language for enhancing AI's legibility.

4. Sign Language for AI

Given the complexity of the issues which confound AI legibility we conceptualised our design challenge as developing a visual 'language' for AI, made up of individual modules which can be combined to develop meaning. The design process we describe here broadly falls into four phases. First, drawing upon prior AI research we identified several key concepts that are relevant to AI legibility. Second, we explored how those concepts may be represented in three different visual styles; a pictorial style, a textual style and an abstract style (see figure

2). Third, we focused on a single one of the styles (the abstract style) and iteratively redesigning icons to develop the core concepts of the visual language. Fourth, inspired by 'Design Fiction' (cf. P. Coulton, Lindley, Sturdee, & Stead, 2017), we began to speculate around what regulatory and social changes would be necessary for widespread and sensible adoption of the icons. Figure 2 describes each of the key AI concepts which we incorporated into the visual language and shows how each one manifested in terms of the three visual styles.



Presence
Examples illustrate that AI is present

The basic form of each icon is meant to represent that some form of AI processing is happening. These icons could be used in software to denote features which use AI, on devices to show that a particular device uses AI, or even in a public space to denote that AI is being used in this area.

Processing Location
Examples illustrate cloud-based AI processing

Representing whether 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 (Rader, Cotter, & Cho, 2018).

Learning Scope
Examples show 2nd Order Adaptive
(eg. AI inputs and outputs to another AI)

Is the AI static, does it adapt based on usage, or is does it behave as a '2nd order adaptive' system (e.g. is fed by, or feeds into, another AI). To involve users in meaningful 'conversation' (Weld & Bansal, 2019) requires explaining whether the AI will adapt based on its use and whether that may impact some other 3rd party system. Communciating how a system will adapt over time is a key guideline for designing AI systems (Amershi et al., 2019).

Data Provenance
Examples show the AI is trained on public data

What is the source of the training data—proprietary, public, or the user themselves? The data quality which goes into an AI often directly reflect its behaviour and are demonstrably significant factor of users' ability to make ethical and moral judgements about an AI (Arnold et al., 2018; Mortier, Haddadi, Henderson, McAuley, & Crowcroft, 2014; Rader et al., 2018).

Training Data Types
Examples show that this AI feature was trained using location data

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 (Burrell, 2016).

Intrinsic Labor
Representation of intrinsic work done for the AI operator by the AI user

By use of the AI feature is 'work' being done for the AI operator? This factor is different to the others in that it is inherently subjective. Reflecting on how the monetization of data is driving the commodification of users and their everyday interactions (cf. Morozov, 2013) this concept strives to communicate the monetary value of users' interactions with an AI, for the AI operator.

Figure 2. Key concepts for visual AI language.

Whilst the AI concepts we chose to work with could never be an exhaustive account of salient AI issues as shown in figure 2, each concept directly relates and contributes to an ongoing area of AI research; accountability (Rader et al., 2018), transparent adaptation (Amershi et al., 2019; Weld & Bansal, 2019), data bias and quality (Arnold et al., 2018; Burrell, 2016; Mortier, Haddadi, Henderson, McAuley, & Crowcroft, 2014), and broader issues relating to social agency and power (Morozov, 2013).

For each concept, figure 2 shows three design approaches. The first (pictorial) design uses a familiar trope of AI iconography—a brain depicted as a network. While clearly problematic in terms of upholding the issues associated with AI's dualism, the brain motif is a symbolic sign and therefore effortlessly carries some (limited) meaning. The second (textual) design employs typography, and whilst also symbolic (i.e. it has no intrinsic meaning), we deliberately combined a branding element (e.g. the 'AI' symbol) with a more communicative element (e.g. 'cloud-based' AI, see figure 2b). When adopted such imagery (e.g. *Fairtrade*, the *Conformité Européenne*—or CE—safety mark) it can become a powerful element of behaviour change (Blythe & Johnson, 2018). The third (abstract) style draws on a design language which hybridises symbolic, indexical, and iconic signs. Whilst some element of convention is necessary to understand these abstract signs, once the core elements of the language are understood this approach has the potential to be interpreted meaningfully (per indexical or iconic signs). For example, if the reader knows that a small dot represents the AI, then a small dot *outside* the icon represents remote, or cloud-based AI, whereas inside would represent local or edge-based AI (see figure 2b).

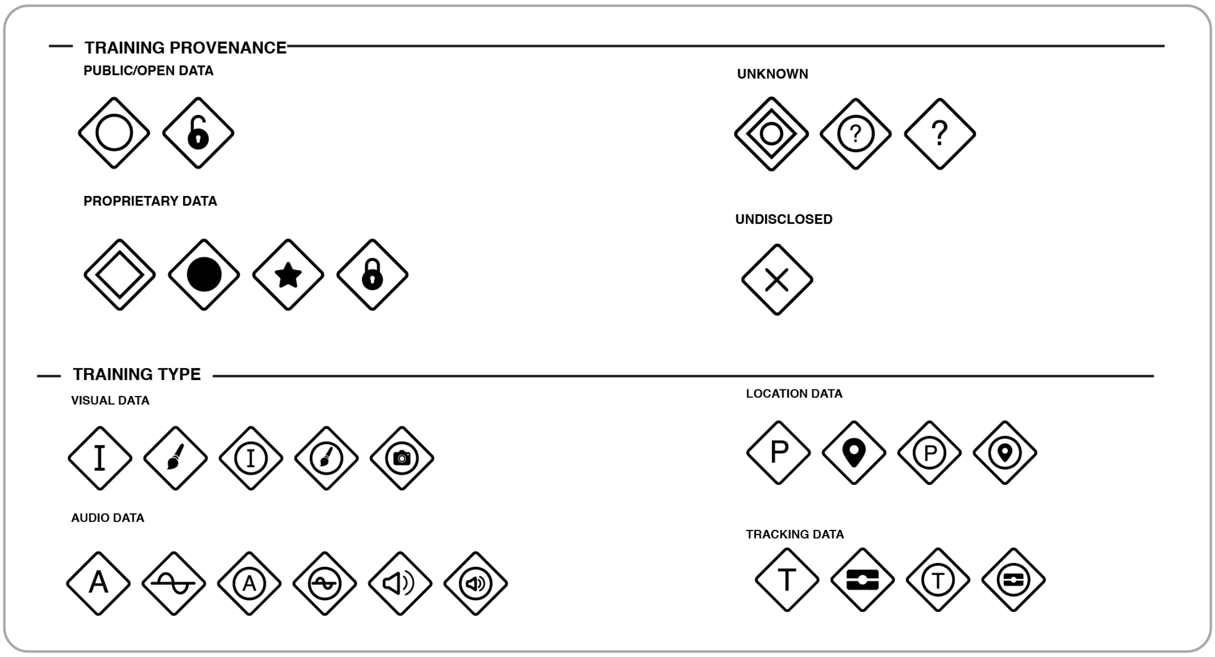
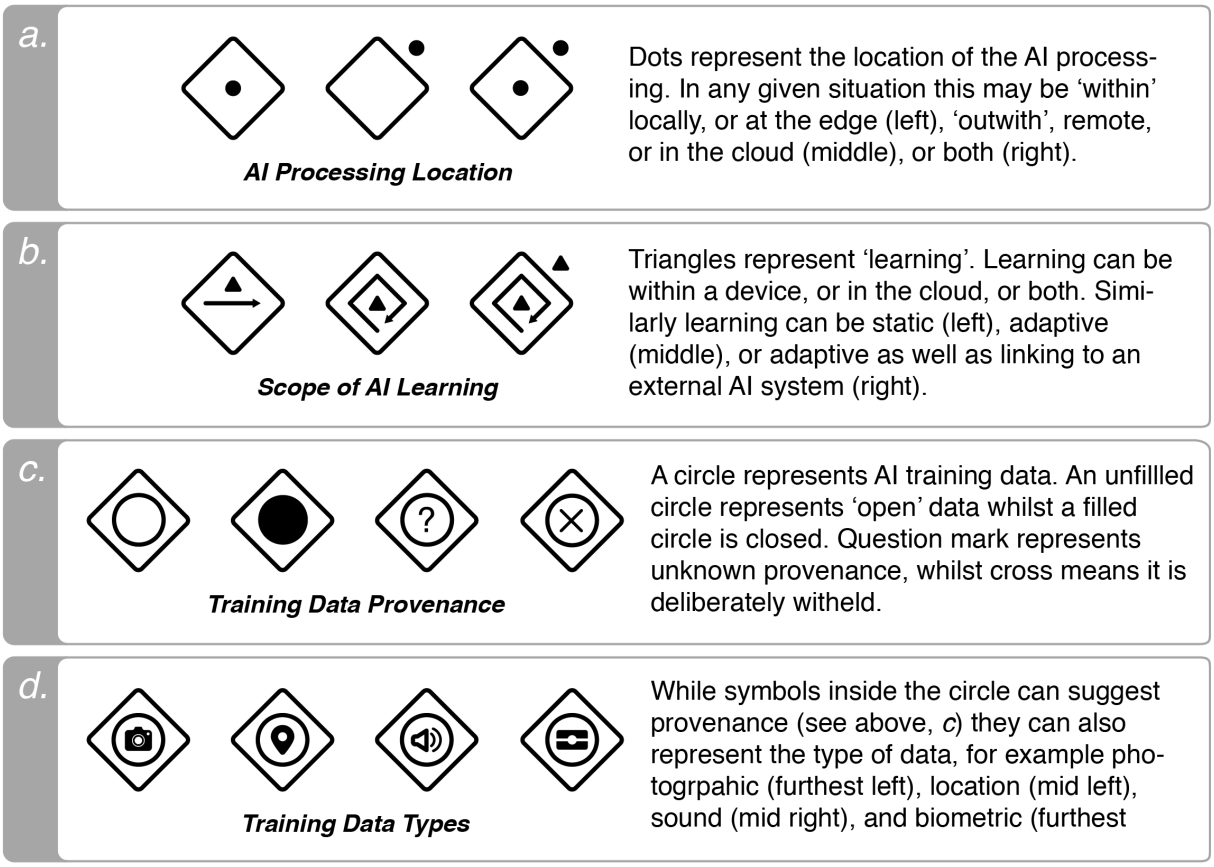


Figure 3. Design iterations of the abstract icon style.

A further exploration of the first two styles will become part of a future co-design research project, but for the purposes of this RtD study we elected to iterate and further develop the third style as it offered the most scope to explore the boundaries of what a visual language for AI could offer. The abstract style offers a unique flexibility, allowing a combination of indexical, iconic, and symbolic elements. The most challenging issue with this approach is how to construct a meaningful grammar about salient elements of AI such as the relationship between data sources, data types, training types, and outcomes. The final iterations of our designs began to address these factors by developing key aspects of a language. For example, ‘dots’ represent AI processing (figure 4a), ‘triangles’ represent AI learning (figure 4b), and relationships with data are denoted by an icon inside a circle (figure 4c, 4d).



Dots represent the location of the AI processing. In any given situation this may be ‘within’ locally, or at the edge (left), ‘outwith’, remote, or in the cloud (middle), or both (right).

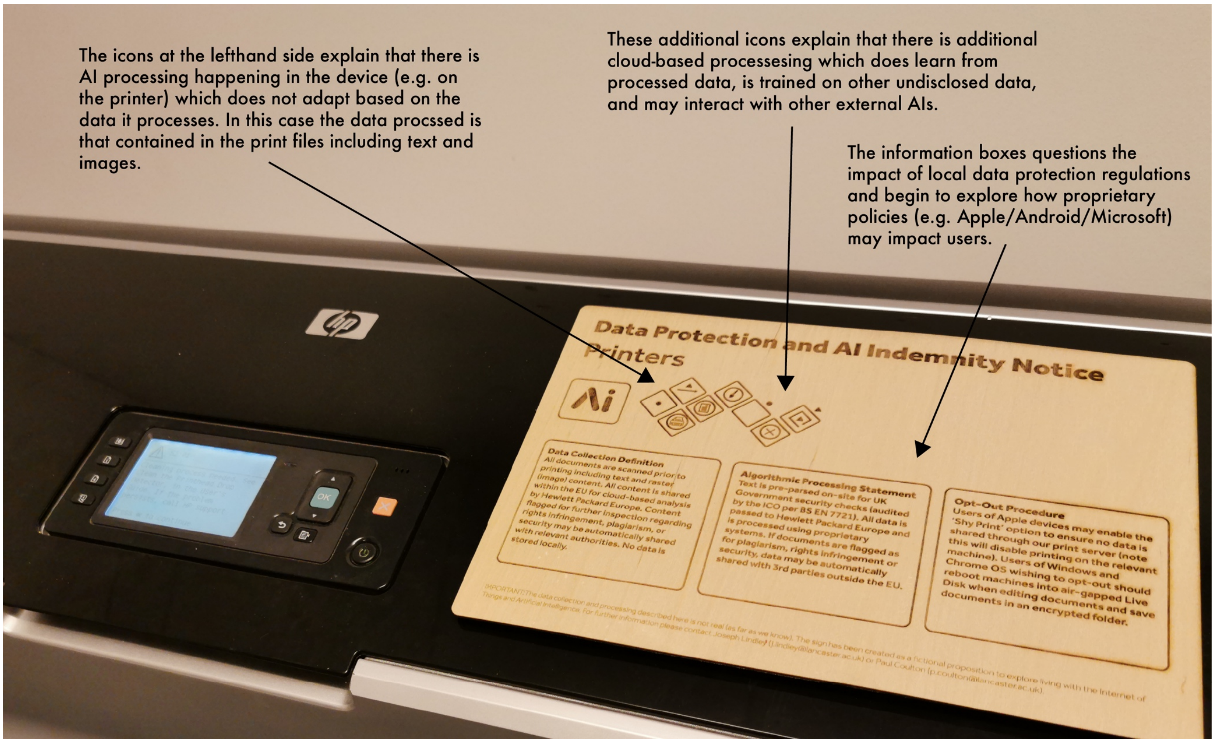
Triangles represent ‘learning’. Learning can be within a device, or in the cloud, or both. Similarly learning can be static (left), adaptive (middle), or adaptive as well as linking to an external AI system (right).

A circle represents AI training data. An unfillled circle represents ‘open’ data whilst a filled circle is closed. Question mark represents unknown provenance, whilst cross means it is deliberately witheld.

While symbols inside the circle can suggest provenance (see above, *c*) they can also represent the type of data, for example photogrpahic (furthest left), location (mid left), sound (mid right), and biometric (furthest

Figure 4. Final iterations of the abstract icon style.

While a study to assess the icons intuitiveness (e.g. Ferreira et al., 2006) and evaluate different designs will produce useful insights and is planned for future work, in this paper we are interested to explore how AI icons may be utilised through practical applications. In order to do this, we employ Design Fiction as World Building. This approach utilises speculative designs as ‘entry points’ into designed imaginaries (P. Coulton et al., 2017). In this case we build upon the icons discussed thus far and incorporated them into information signs intended to show employees and visitors, in a workplace environment, how AI is used in several mundane contexts; for printing (figure 5), going to the toilet (figure 6), in a computer suite (figure 7), and with security cameras (figure 8). In the speculative world the signs were designed assuming that AI is used ubiquitously and that conventions have been established insisting that uses of AI are signposted in public—these attributes are alluded to in the signs themselves. Each sign follows the same layout. They are cast as generic ‘Data Protection and AI Indemnity Notices’. Each one incorporates configurations of our icon designs, used modularly to describe how that service utilises and interacts with AI. Three text-based elements on each sign describe how the AI services use data, what sort of processing takes place, and how users might opt out of the AI altogether.



The icons at the lefthand side explain that there is AI processing happening in the device (e.g. on the printer) which does not adapt based on the data it processes. In this case the data procssed is that contained in the print files including text and images.

These additional icons explain that there is additional cloud-based processesing which does learn from processed data, is trained on other undisclosed data, and may interact with other external AIs.

The information boxes questions the impact of local data protection regulations and begin to explore how proprietary policies (e.g. Apple/Android/Microsoft) may impact users.

Data Protection and AI Indemnity Notice
Printers

Data Collection Definition
All documents are scanned prior to printing including text and raster (image) content. All content is shared within the EU for cloud-based analysis by Hewlett Packard Europe. Content flagged for further inspection regarding rights infringement, plagiarism, or security may be automatically shared with relevant authorities. No data is stored locally.

Algorithmic Processing Statement
Text is pre-parsed on-site for UK Government security checks (audited by the ICO per BS EN 7721). All data is passed to Hewlett Packard Europe and is processed using proprietary systems. If documents are flagged as for plagiarism, rights infringement or security, data may be automatically shared with 3rd parties outside the EU.

Opt-Out Procedure
Users of Apple devices may enable the ‘Shy Print’ option to ensure no data is shared through our print server (note this will disable printing on the relevant machine). Users of Windows and Chrome OS wishing to opt-out should reboot machines into air-gapped Live Disk when editing documents and save documents in an encrypted folder.

IMPORTANT: The data collection and processing described here is not real (as far as we know). The sign has been created as a fictional proposition to explore living with the Internet of Things and Artificial Intelligence. For further information please contact Joseph Lindley (j.lindley@lancaster.ac.uk) or Paul Coulton (p.coulton@lancaster.ac.uk).

Figure 5. Sign attached to a printer utilising AI for various purposes including plagiarism checking and copyright protection.

This printer depicted in Figure 5 evidently sends *all* data that it will eventually print for processing by the manufacturer in order to be checked for copyright infringements. The notice informs us that *no* data is stored locally. A security check is also run before the data is passed, which appears to regulated by the Information Commissioners Office and a European standard, if content is flagged for further security or rights checks then it will be shared outside of the EU (and presumably would not then be protected by the EU’s data protection legislation). The opt-out section demonstrates how different manufacturers can choose to implement the ability to opt-out differently, in this case Apple has made an easy to use feature (but if users employ it then they can no longer print), such provisions are not so accessible for users of other operating systems.

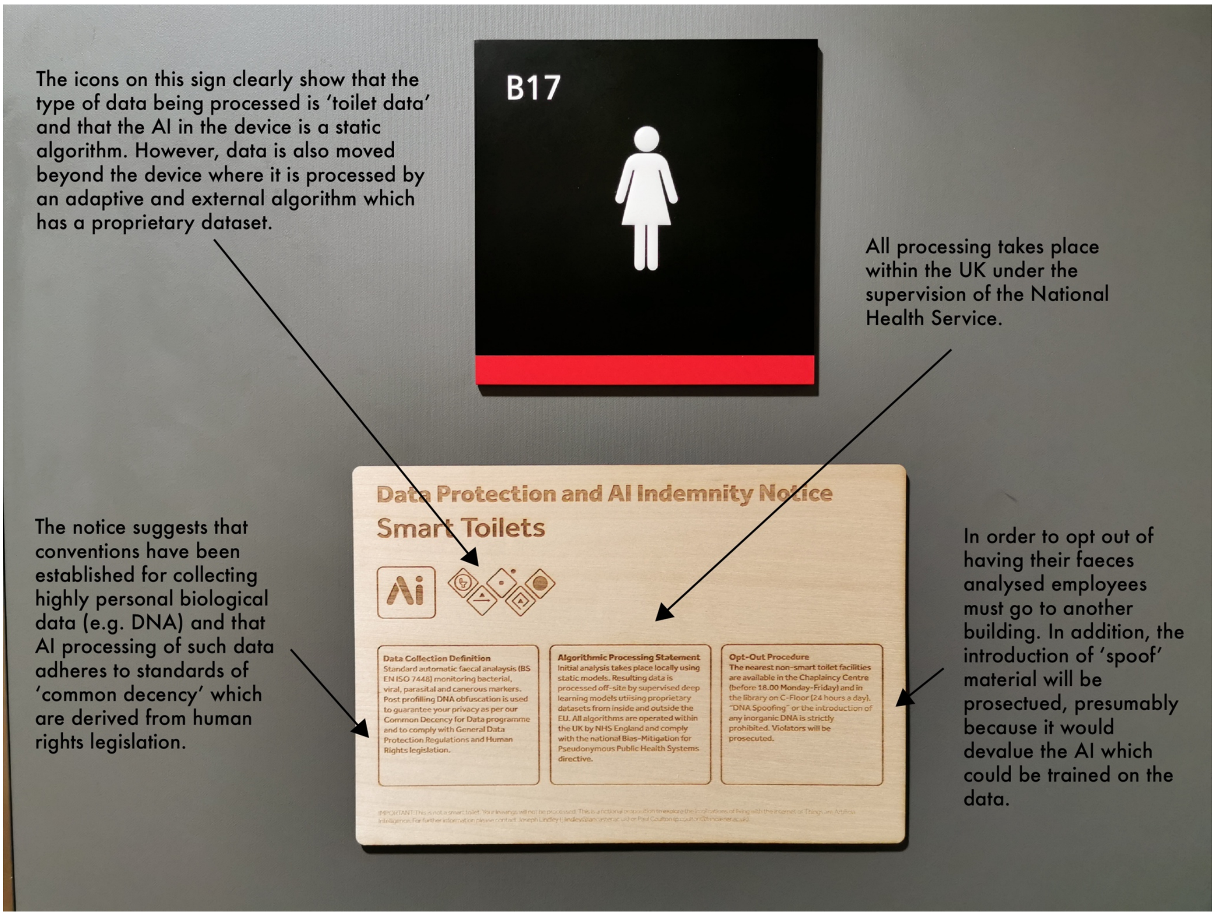
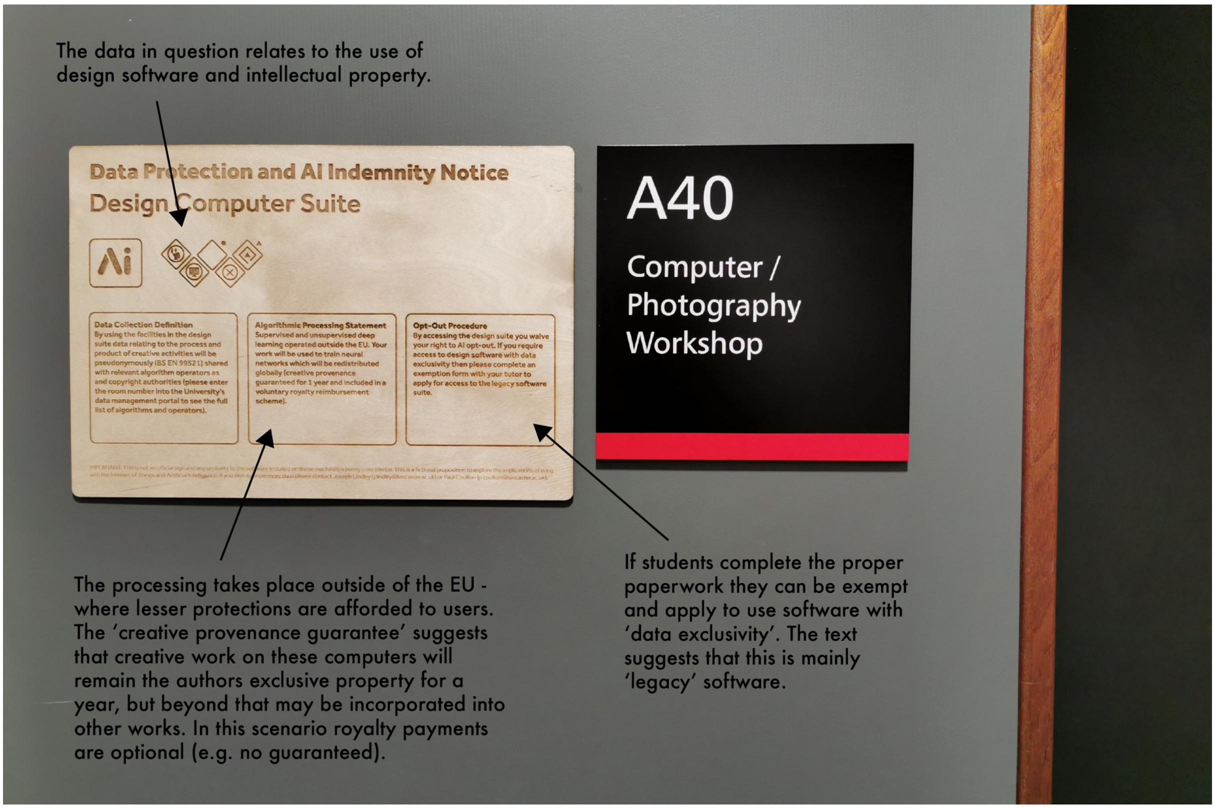

The icons on this sign clearly show that the type of data being processed is 'toilet data' and that the AI in the device is a static algorithm. However, data is also moved beyond the device where it is processed by an adaptive and external algorithm which has a proprietary dataset.
B17
All processing takes place within the UK under the supervision of the National Health Service.
Data Protection and AI Indemnity Notice
Smart Toilets
Ai
The notice suggests that conventions have been established for collecting highly personal biological data (e.g. DNA) and that AI processing of such data adheres to standards of 'common decency' which are derived from human rights legislation.
Data Collection Definition
Standard automatic faecal analaysis (BS EN ISO 7448) monitoring bacterial, viral, parasital and canerous markers. Post profilling DNA obfuscation is used to guarantee your privacy as per our Common Decency for Data programme and to comply with General Data Protection Regulations and Human Rights legislation.
Algorithmic Processing Statement
Initial analysis takes place locally using static models. Resulting data is processed off-site by supervised deep learning models utiising proprietary datasets from inside and outside the EU. All algorithms are operated within the UK by NHS England and comply with the national Bias-Mitigation for Pseudonymous Public Health Systems directive.
Opt-Out Procedure
The nearest non-smart toilet facilities are available in the Chaplaincy Centre (before 18.00 Monday-Friday) and in the library on C-Floor (24 hours a day). "DNA Spoofing" or the introduction of any inorganic DNA is strictly prohibited. Violators will be prosecuted.
In order to opt out of having their faeces analysed employees must go to another building. In addition, the introduction of 'spoof' material will be prosectued, presumably because it would devalue the AI which could be trained on the data.
IMPORTANT: This is not a smart toilet. Your leavings will not be processed. This is a fictional proposition to explore the implications of living with the Internet of Things and Artificial Intelligence. For further information please contact Joseph Lindley (j.lindley@lancaster.ac.uk) or Paul Coulton (p.coulton@lancaster.ac.uk).

Figure 6. Sign on the entrance to toilets equipped with 'smart toilet' AI processing.

Figure 6 explores the highly personal concept of using AI to analyse DNA, as well as 'bacterial, viral, parasitical and cancerous' markers. The sign suggests that, although opting out is a possibility (protected by a legal right) it is practically hard to achieve for those wishing to use the toilet. The icons show that the primary processing not adaptive (i.e. the system's processing not adapt based on your leavings), however, it also shares data beyond which may be used in proprietary data and AI systems. This configuration may reassure users who consider the 'headline' icons (that the processing is linear and local) but could confuse others that continue to read the icons. It raises the need to explore how a 'grammar' may be used to provide the icons with more semantic nuance.



The data in question relates to the use of design software and intellectual property.

Data Protection and AI Indemnity Notice

Design Computer Suite

Data Collection Definition
By using the facilities in the design suite data relating to the process and product of creative activities will be pseudonymously (BS EN 99521) shared with relevant algorithm operators as and copyright authorities (please enter the room number into the University's data management portal to see the full list of algorithms and operators).

Algorithmic Processing Statement
Supervised and unsupervised deep learning operated outside the EU. Your work will be used to train neural networks which will be redistributed globally (creative provenance guaranteed for 1 year and included in a voluntary royalty reimbursement scheme).

Opt-Out Procedure
By accessing the design suite you waive your right to AI opt-out. If you require access to design software with data exclusivity then please complete an exemption form with your tutor to apply for access to the legacy software suite.

IMPORTANT: This is not an official sign and any similarity to the software installed on these machines is purely coincidental. This is a fictional proposition to explore the implications of living with the Internet of Things and Artificial Intelligence. If you wish to know more then please contact Joseph Lindley (j.lindley@lancaster.ac.uk) or Paul Coulton (p.coulton@lancaster.ac.uk).

A40

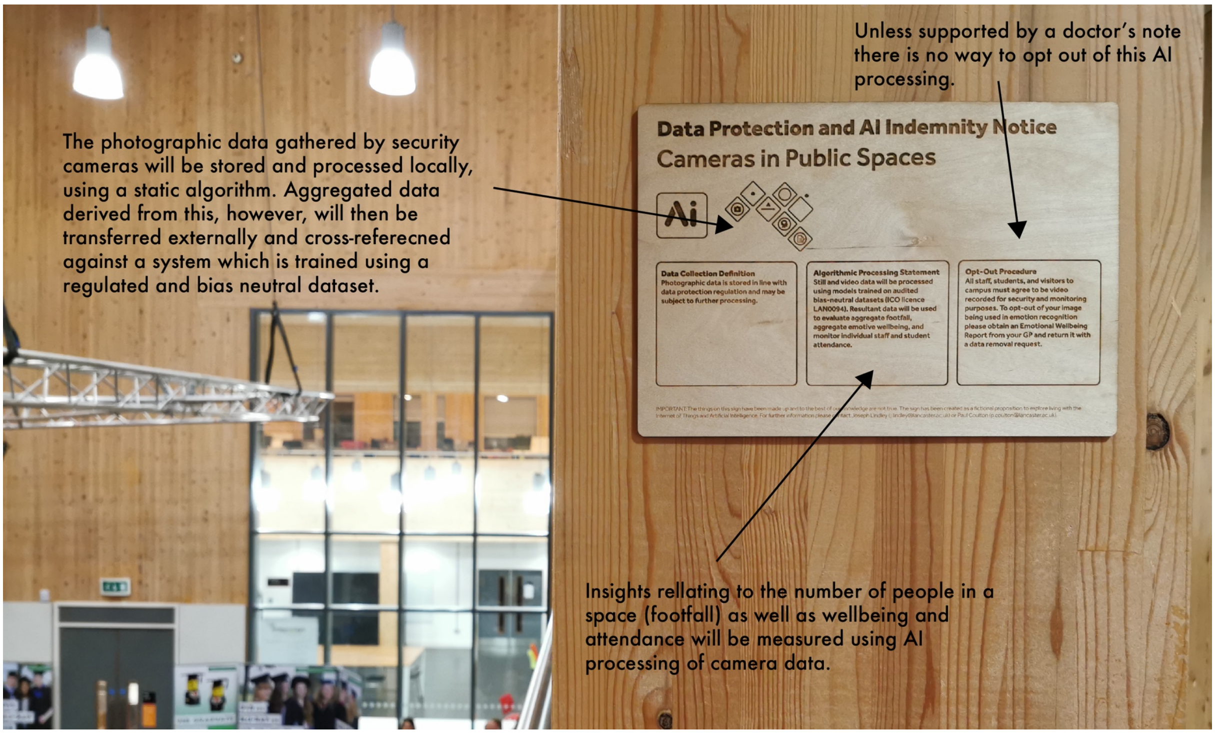
Computer / Photography Workshop

The processing takes place outside of the EU - where lesser protections are afforded to users. The 'creative provenance guarantee' suggests that creative work on these computers will remain the authors exclusive property for a year, but beyond that may be incorporated into other works. In this scenario royalty payments are optional (e.g. no guaranteed).

If students complete the proper paperwork they can be exempt and apply to use software with 'data exclusivity'. The text suggests that this is mainly 'legacy' software.

Figure 7. Creative software packages, such as those used in the photography workshop, may be subject to AI processing as well.

The relationship between AI and creativity will inevitably raise dilemmas around authorship and ownership. If an AI aids a creator in their work, does it deserve some form of credit? Conversely, if creative work is used to train an AI which aids in creative work, then should subsequent works which utilise that AI credit the authors of the training works? In this example the icons show how any work done in the computing suite will be subject to external processing of any 'intellectual' data, outside of any EU protection, and with no guarantees that derivative works will credit the original creator. As AI becomes more central to creative workflows, challenges around provenance and authenticity will grow. Our example signs are set in a University context; the dynamic between students' creative acts and intellectual property in Figure 7 raises the questions the prestige of institutions and the types of AI they employ. Would wealthier, higher performing, Universities have access to better and less intrusive AI systems?



Unless supported by a doctor's note there is no way to opt out of this AI processing.

The photographic data gathered by security cameras will be stored and processed locally, using a static algorithm. Aggregated data derived from this, however, will then be transferred externally and cross-referecned against a system which is trained using a regulated and bias neutral dataset.

Data Protection and AI Indemnity Notice

Cameras in Public Spaces

Data Collection Definition
Photographic data is stored in line with data protection regulation and may be subject to further processing.

Algorithmic Processing Statement
Still and video data will be processed using models trained on audited bias-neutral datasets (ICO licence LAN0094). Resultant data will be used to evaluate aggregate footfall, aggregate emotive wellbeing, and monitor individual staff and student attendance.

Opt-Out Procedure
All staff, students, and visitors to campus must agree to be video recorded for security and monitoring purposes. To opt-out of your image being used in emotion recognition please obtain an Emotional Wellbeing Report from your GP and return it with a data removal request.

IMPORTANT: The things on this sign have been made up and to the best of our knowledge are not true. The sign has been created as a fictional proposition to explore living with the Internet of Things and Artificial Intelligence. For further information please contact Joseph Lindley (j.lindley@lancaster.ac.uk) or Paul Coulton (p.coulton@lancaster.ac.uk).

Insights rellating to the number of people in a space (footfall) as well as wellbeing and attendance will be measured using AI processing of camera data.

Figure 8. Security camera footage may be processed by AI in order to derive a range of insights.

The ubiquitous capturing and recording of our image is a fact of modern life. Mainly for security reasons most public spaces are under the purview of digital cameras. At the present time, for the most part, these images are simply archived for later perusal if needed. AI technology, however, would provide practicable means to utilise such images in more complex ways. From simply detecting how many people are in a space, through to understanding who is in the space, and the emotional state of those individuals—the array of insights that AI processing of photographic imagery can produce is immense. The sign in Figure 8 suggests that this kind of monitoring is not optional for employees of the organisation in question, or guests attending the building. The only occasion that an employee may apply to opt-out is if they have a doctor's note showing that the surveillance is doing them harm. Whilst there is clearly potential for positive uses of this data—e.g. to optimise the use of space and maximise employee wellbeing—such aspirations must be balanced against a sovereignty of the self. Although our relationship with cameras in public spaces is generally one of reassurance, as AI processing becomes viable, we may need to reassess how we make judgements about our image being captured.

5. Discussion and Future Work

As is common with RtD-based inquiry, this work aspires to produce contingent findings (Gaver, 2012). The spaces that RtD is adept at studying—in this case AI legibility—tend to be in flux, and hence any findings, whilst aspirational, should be interpreted relative to that flux. That is not to say that such findings are not useful, but rather that they are subject to ongoing interpretations—we suggest that the work presented here should be seen on those terms. Moreover, the principal weight of the research is not carried solely within the designed outcomes, the process, or related literature. The insights, in fact, emerge from considering all three of these aspects at the same time, and the remainder of the discussion reflects on each of them. To that end we reiterate that the target audience for the designs presented in the paper are the AI and Design research communities. Whilst we are confident that there is some merit, in some of the signs, in their current form they are not intended for a general audience, but rather to being the significant task of utilising Design Research to strive towards legible, and responsible, AI innovation.

The volume and diversity of research into AI is representative of its existing and future impact on the world. However, the landscape is unbalanced. Whilst applications of AI continue to be adopted at a pace—largely driven by the private sector—the extensive efforts to develop frameworks, taxonomies, and social standards for the understanding and acceptance of AI a foundering. The difficulties around public perception of AI, which we cast in this paper as it's definitional dualism, confound this challenge. Design Research has key roles to play in both unifying aspects of disparate perspectives (e.g. synthesising both technical and social research) and also framing AI rhetoric in such a way that it reflects gravity and scope of AI adoption. Whilst the prevailing rhetoric places AI as a proximate future, in reality AI is here *now*. To that end, via both the desk-based and practice-based elements of this paper we hope that the unique and important role for design research has been highlighted.

This is early stage work, yet the process of designing and developing the icons and the signs has helped develop a range of insights which apply at various scales. The technologies depicted are all being actively developed, and as such the focal point of our enquiry was not so much the technologies themselves, but the reality which they exist within. In our reality, although the use of AI is quite intrusive (e.g. analysing faecal matter or scanning printed documents) the intrusion is conducted within a strict regulatory environment. The multiple authorities involved (e.g. information commissioners, standards organisations) and inter-related policies (e.g. healthcare legislation, data protection law, local organisational policies) make the otherwise intrusive use of AI seem more innocuous, however the practicalities of implementing such a complex regulatory environment are not insignificant. Similarly, an assumption within all of the signage we created is some kind of agreement, or didactic ruling, about what AI actually is—what classes of AI need to be regulated, and in what contexts? Should AI-based processing of printed documents in a workplace be held to the same standard as analysing employees' poo? In addition to the broader context that AI signs exist within issues relating to the minutiae of the problem also arose. While we focused on a modular and abstract icon design, the food industry demonstrates a huge variety of iconographic ways of communicating about the product; emblems and logos tell us whether food is organic, fair trade, vegan, all natural, high in protein, etc. Prior research suggests that in some circumstances the very presence of these signs reduces critical engagement with the issues the sign is addressing (Blythe & Johnson, 2018). Supported by our own experience of attempting to craft signs that were legible, we considered that even if a mandate for public signage existed, establishing if signs are efficacious may be a complex and ongoing task. To draw upon the terminology tactfully offered by one reviewer of the paper, even if the signs are carriers of information how do we know if they are carriers of meaning—establishing this is difficult and may call upon a collaboration between design and other research communities. It would, perhaps, be through such a process that iterations of a visual language for AI legibility would move from being purely a research instrument to a viable or implementable product[3].

This research does not aspire to provide definitive answers to explicitly defined research questions, but rather provide contingent insights relating to the ongoing impact of AI's adoption. The contributions of this paper are multiple. First, by reviewing a range of AI-related literature we highlight the cross-sectoral and multi-disciplinary challenges that AI poses. Next we introduce the crucial integrative role of design-led research can play by making aspects of other research programmes tangible and providing a sensible framework to reflect on them. Finally, though the reflexive process of designing icons and signage aimed at AI legibility we begin to frame questions and pathways for future research. Ongoing work in this area must be multifaceted. Clear avenues include empirical assessments and iterative developments of visual cues to support AI legibility. Further speculative design work, incorporating participatory and co-designed aspects, will develop practical means to integrate AI research, helping create coherent research programmes out of disparate research projects and in doing so, help to develop research instruments commensurate with the challenges posted by AI.

Acknowledgements: We would like to express our gratitude towards the anonymous reviewers who provided invaluable insights which have been incorporated into the paper. This work was supported the UK Engineering and Physical Sciences Research Council under grant reference EP/S035362/1. Finally, thanks to our colleagues for occasionally encountering signs relating to fictional AI systems and for not panicking!

5. References

Amershi, S., Chickering, M., Drucker, S. M., Lee, B., Simard, P., & Suh, J. (2015). ModelTracker. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (pp. 337–346). New York, New York, USA: ACM Press. https://doi.org/10.1145/2702123.2702509

Amershi, S., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E., Weld, D., … Bennett, P. N. (2019). Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19* (pp. 1–13). New York, New York, USA: ACM Press. https://doi.org/10.1145/3290605.3300233

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal

[3] Note that, as eluded to by one of our reviewers, were such an implementation to be pursued cultural nuances vis-à-vis semiotics—which are currently entirely omitted from this study—must be considered.

and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Arnold, M., Bellamy, R. K. E., Hind, M., Houde, S., Mehta, S., Mojsilovic, A., … Varshney, K. R. (2018). FactSheets: Increasing Trust in AI Services through Supplier’s Declarations of Conformity. Retrieved from http://arxiv.org/abs/1808.07261

Blythe, J., & Johnson, S. (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

Buchanan, R. (1992). Wicked Problems in Design Thinking. *Design Issues*, *8*(2), 5–21.

Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 205395171562251. https://doi.org/10.1177/2053951715622512

Cave, S., & ÓhÉigeartaigh, S. S. (2018). An AI Race for Strategic Advantage. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society - AIES ’18* (pp. 36–40). New York, New York, USA: ACM Press. https://doi.org/10.1145/3278721.3278780

Cooper, R., Dunn, N., Coulton, P., Walker, S., Rodgers, P., Cruikshank, L., … Coulton, C. (2018). ImaginationLancaster: Open-Ended, Anti-Disciplinary, Diverse. *She Ji: The Journal of Design, Economics, and Innovation*, *4*(4), 307–341. https://doi.org/10.1016/j.sheji.2018.11.001

Coulton, P., & Lindley, J. G. (2019). More-Than Human Centred Design: Considering Other Things. *The Design Journal*, 1–19. https://doi.org/10.1080/14606925.2019.1614320

Coulton, P., Lindley, J., Sturdee, M., & Stead, M. M. (2017). Design Fiction as World Building. In *Proceedings of the 3rd Biennial Research Through Design Conference* (pp. 1–16). Edinburgh, UK. https://doi.org/10.6084/m9.figshare.4746964.v2

Ferreira, J., Barr, P., & Noble, J. (2002). The Semiotics of User Interface Redesign. In *Proceedings of the Sixth Australasian conference on User interface* (Vol. 40, pp. 47–53).

Ferreira, J., Noble, J., & Biddle, R. (2006). A case for iconic icons. In *Conferences in Research and Practice in Information Technology Series* (Vol. 50, pp. 87–90).

Frayling, C. (1993). Research in Art and Design. *Royal College of Art Research Papers*, *1*(1), 1–9.

Gaver, W. (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* (pp. 937–946). New York, New York, USA: ACM Press. https://doi.org/10.1145/2207676.2208538

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

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

Haddadi, H., Mortier, R., McAuley, D., & Crowcroft, J. (2012). *Human Data-Interaction*.

Hall, W., & Pesenti, J. (2017). *Growing the Artificial Intelligence Industry in the Uk*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf

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

Industrial Strategy. (2017). Retrieved from https://www.gov.uk/government/topical-events/the-uks-industrial-strategy

Kiran, A. H., & Verbeek, P.-P. (2010). Trusting Our Selves to Technology. *Knowledge, Technology & Policy*, *23*(3–4), 409–427. https://doi.org/10.1007/s12130-010-9123-7

Lanier, J. (2013). *Who owns the future*. Simon and Schuster.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 205395171875668. https://doi.org/10.1177/2053951718756684

Lindley, J., Akmal, H. A., & Coulton, P. (2020). Design Research and Object-Oriented Ontology. *Open Philosophy*, *3*(1), 11–41. https://doi.org/10.1515/opphil-2020-0002

Lindley, J., Coulton, P., Akmal, H. A., Hay, D., Van Kleek, M., Cannizzaro, S., & Binns, R. (2019). Fairness. In C. Coulton (Ed.), *Little Book of Philosophy for the Internet of Things*. ImaginationLancaster. Retrieved from https://www.petrashub.org/download/the-little-book-of-philosophy-for-the-internet-of-things/

Lindley, J., Coulton, P., & Sturdee, M. (2017). Implications for Adoption. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17* (pp. 265–277). New York, New York, USA: ACM Press. https://doi.org/10.1145/3025453.3025742

Ma, X., Matta, N., Cahier, J.-P., Qin, C., & Cheng, Y. (2015). From action icon to knowledge icon: Objective-oriented icon taxonomy in computer science. *Displays*, *39*, 68–79. https://doi.org/10.1016/j.displa.2015.08.006

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

Mortier, R., Haddadi, H., Henderson, T., McAuley, D., & Crowcroft, J. (2014). Human-Data Interaction: The Human Face of the Data-Driven Society. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2508051

Rader, E., Cotter, K., & Cho, J. (2018). Explanations as Mechanisms for Supporting Algorithmic Transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (Vol. 2018-April, pp. 1–13). New York, New York, USA: ACM Press. https://doi.org/10.1145/3173574.3173677

Redström, J., & Wiltse, H. (2019). Changing Things: Innovation through Design Philosophy. In *Proceedings of the Academy for Design Innovation Management Conference*.

Ribeiro, M. T., Singh, S., & Guestrin, C. (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* (pp. 1135–1144). New York, New York, USA: ACM Press. https://doi.org/10.1145/2939672.2939778

Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Weld, D. S., & Bansal, G. (2019). The challenge of crafting intelligible intelligence. *Communications of the ACM*, *62*(6), 70–79. https://doi.org/10.1145/3282486

About the Authors:

Joseph Lindley is a Research Fellow interested in how Design Research can contribute towards radical-yet-responsible applications of contemporary technologies including Artificial Intelligence and the Internet of Things.

Haider Ali Akmal is a doctoral candidate and practicing visual artist. His research focuses on the intricacies of human experience as it

relates to digital and physical interactions. His work combines Speculative Design, Philosophy, and Play as instruments of research.

Franziska Pilling is doctoral candidate with a research focus on Artificial Intelligence as a Material for Design and employs Speculative Design, Design Fiction and Experience Design to research alternative perspectives on Artificial Intelligence.

Paul Coulton holds a chair in Speculative and Game Design at Imagination Lancaster. Having allegorically achieved his goal of becoming the first Professor Applied Triviality he is curating a ‘Caravan of the Future’ in order to deliver research impact in hard to reach communities.