Manchester Metropolitan University

e-space

Manchester Metropolitan University's Research Repository

Please cite the Published Version

Pillling, Franziska, Akmal, Haider, Coulton, Paul and Lindley, Joseph (2020) The process of gaining an AI Legibility Mark. In: CHI '20: CHI Conference on Human Factors in Computing Systems, 25 April 2020 - 30 April 2020, Honolulu, Hawaii, USA.

DOI: https://doi.org/10.1145/3334480.3381820

Publisher: Association for Computing Machinery (ACM)

Version: Accepted Version

Downloaded from: https://e-space.mmu.ac.uk/633823/

Usage rights: © In Copyright

Additional Information: © ACM, 2020. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in CHI'20 Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts http://doi.acm.org/10.1145/3334480.3381820

Enquiries:
If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines)

The process of gaining an AI Legibility mark

Franziska Pilling
Lancaster University
Lancaster, UK
f.pilling@lancaster.ac.uk

Joseph Lindley
Lancaster University
Lancaster, UK
j.lindley@lancaster.ac.uk

Haider Akmal
Lancaster University,
Lancaster, UK
h.a.akmal@lancaster.ac.uk

Paul Coulton
Lancaster University,
Lancaster, UK
p.coulton@lancaster.ac.uk

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CHI'20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-6819-3/20/04…$15.00
https://doi.org/10.1145/3334480.3381820

Abstract
Researchers and designers working in industrial sectors seeking to incorporate Artificial Intelligence (AI) technology, will be aware of the emerging International Organisation for AI Legibility (IOAIL). IOAIL was established to overcome the eruption of obscure AI technology. One of the primary goals of IOAIL is the development of a proficient certification body providing knowledge to users regarding the AI technology they are being exposed to. To this end IOAIL produced a system of standardised icons for attaching to products and systems to indicate both the presence of AI and to increase the legibility of that AI's attributes. Whilst the process of certification is voluntary it is becoming a mark of trust, enhancing the usability and acceptability of AI-infused products through improved legibility. In this paper we present our experience of seeking certification for a locally implemented AI security system, highlighting the issues generated for those seeking to adopt such certification.

Author Keywords
Artificial Intelligence; Legibility; Marks; Transparency; Design Fiction.

CSS Concepts
• Computing methodologies ~ Discourse, dialogue and pragmatics.

Introduction

Technology and humanity have co-existed, dependent on each other, for millennia. While operating most forms of tech requires the presence of some form of intelligence, implementing intelligence within technology has only recently become a significant concern. The predominance of AI in our everyday lives should not be understated; myriads of AI-powered artefacts have been universally available for use, for some time now, from smart phones to self-lacing shoes. There are countless interfaces present in our surroundings which oftentimes utilise or are affected by AI; everyday life is peppered with AI. A mundane activity such as taking a bus gives an impression of this gamut, for instance: a bank or travel card is used with digital ticketing terminals interacting with not only banks but also a personal device like a phone; the seat being sat on gathers statistical data on passenger numbers which are processed by AI; on-board wireless connectivity provides connectivity for passengers as well as collecting meta-data for AI-based analysis; on-board security cameras collect facial and biometric data; the autonomous vehicle itself is equipped with a wide variety of sensors interacting with AI systems; upon reaching the destination interactive displays at the bus stop present AI-curated advertisements based on the data mined through the course of your interactions on the journey. These multifarious digital interactions often go unnoticed to users due to a lack of transparency of the surrounding technology and AI agents present.

It is no wonder that in a world where our everyday activities are subject to ubiquitous AI processing, that concerns would arise about making these interactions legible. Underpinned by the General Data Protection Regulations (GDPR), galvanised into the Rights for Users of AI Act, the evolution of the International Organization for Artificial Intelligence Legibility (IOAIL) has been deemed necessary for quelling the growing societal concerns associated with the use of AI. The IOAIL logo accompanied by its myriad iconography depicting the operational character of AI present within a device has become a familiar sight for users, as familiar as laundry labels and traffic signs. The long-established laundry labelling scheme, however, highlights how such systems are not always intuitively readable to the end user. IOAIL iconography, while proving successful; in some trials [3], has been cited as "mysterious and bamboozling" [Ibid]. Mystique aside, the requirement to include the IOAIL symbols on artefacts that utilize AI comes from a need for standardisation already seen in different avenues of design and technology development [1, 4, 5].

Market Square Council AI implementation and legibility

Market Square is in the city centre of the Northern UK city of Lancaster. A pedestrian communal hub with eateries, banks, a city library, and occasionally the centre for bustling farmers markets and events. Despite being a popular spot for locals, it has often been the target of both low-level crime, such as bag snatching, and higher-level crimes, such as a recent string of bank robberies. To avoid a drop-in footfall for local businesses, Lancaster council sought funding to implement a state-of-the-art AI security system. This system is one of many uniquely implemented security packages offered by the company AI Security Ops. The particular package in place at Market Square is a comprehensive system with 9 interlinked cameras covering the full extent of the square, with

concentrated views towards high target areas of the square such as cash machines and shops (**Figure 1**). These cameras are AI assisted with gait and facial recognition software connected to a networked crime database. There are also microphones situated in many areas of the square, recording a large amount of data for the purposes of security and raw material for machine learning.

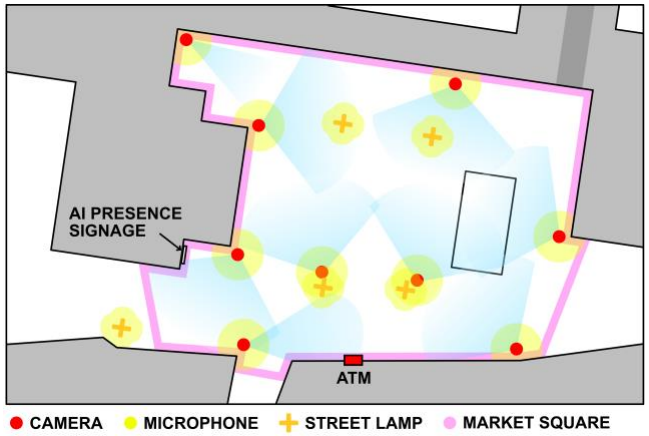

Figure 1: A bird's eye view of Market Square, Lancaster complete with the AI security package in place. This maps the multiple cameras, their angles of view and microphone placement covering the entire square.

Locals of Lancaster were initially pleased by the extra security measures in place in Market Square, providing them with a sense of safety that police would be notified if unusual behaviour was detected by the AI, or biometric readings would spot wanted or known criminals. The security system in place had a positive impact, illustrated by the significant fall in the rate of crime committed in the square. However, as has been previously seen with CCTV systems crime spread to other areas of the city not covered by the AI security, emphasising the technical superiority of the security system in place at Market Square. This disparity brought attention to the AI system and the operation of it to the locals with concerns regarding the lack of transparency of how data was collected and used when crime had relocated to other areas. The only indication that AI was an integral part of the security system was a singular sign the council installed to alert those entering the square that AI enhanced security was present in this location (**Figure 2**).

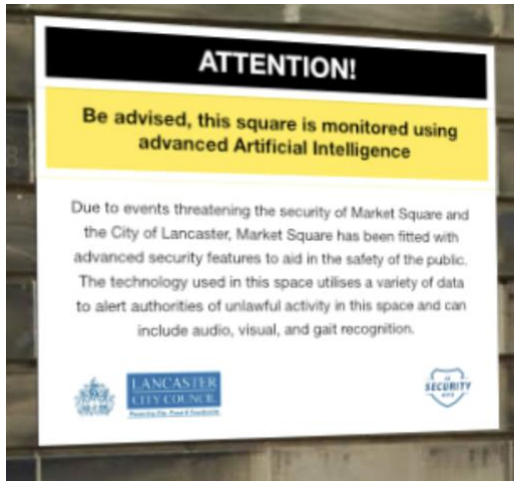

Figure 2: Vague mark of AI presence in Market Square.

Detailing IOAIL and research process

As researchers within PETRAS hub, we too were concerned about the lack of legibility of the AI system in place and the vague indication of its presence, which presented more questions than answers. To this end, we with the council's permission, sought to better understand the process of acquiring a certified mark of

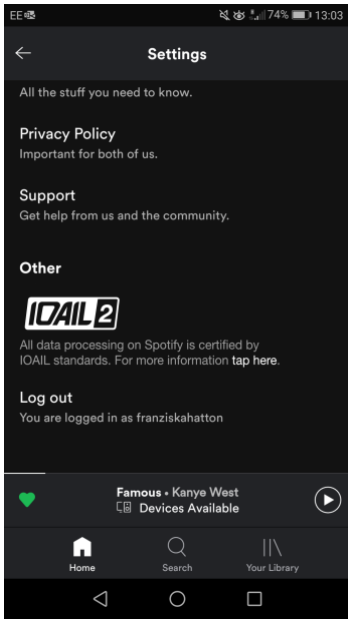

Figure 4: Spotify disclaimer of IOAIL mark can be found in application settings.

AI presence and legibility as provided by the IOAIL, with the hope that the process of marking technology presences and attributes in a public space becomes a routine process for councils and governments.

As previously mentioned, the IOAIL is quickly becoming a prolific certification body with many technology companies noticing that consumers are taking it into account when selecting AI products and services. The lack of an IOAIL mark for many consumers may lead to them questioning the meaning of such absence in regard to how their data is being used and by whom.

The IOAIL has established a visual language of labels and icons to make users' interactions with AI legible developed from historical research into communicating the inner workings of AI technology [7]. As well as icons communicating the functionalities of an AI system for a user, the IOAIL has created an overall class system as a quick indication for users of how much of the AI system is known (**Figure 3**), for instance if a product has a mark of IOAIL 1 than it would be a sign that this product and the owner has not disclosed many attributes of the AI present in the technology being scrutinised, an IOAIL 3 mark is a sign that the product and owners have disclosed all attributes and the AI is effectively legible. Well-known companies that have placed their products through the process of gaining IOAIL marks are: Spotify with an IOAIL 2 Class for their platform (**Figure 4**), and Tesla who certified their latest car the *Model S AI* with an IOAIL 1 Class mark. AI systems within the class spectrum can be at the high-end or low-end of a class though certain attributes might fall in supposed 'grey areas' in between classes. These grey areas arise from insufficient information coming from the technology/service or its creators due to trade secrets [2]. Ergo the holding back of information regarding the AI's capabilities may result in a lower class being assigned. That said, the IOAIL for rigor takes the average a product or system has scored and presents that finding over a spectrum of possible results exhibiting the potential grey areas in the AI that consumers might be interested in looking into.

Figure 3: These marks have been designed in house at IOAIL as a preliminary guide for users to quickly identify the overall transparency and legibility a particular AI infused product has. Further details of these particulars can be found by the user if required.

One particular example, often cited in arguments surrounding the lack of legibility in AI, is the now infamous Roomba-Gate controversy (**Figure 5**). Consisting of Amazon's autonomous vacuum cleaner, the *iRobot Roomba* and the revealing of its sourcing dimensional data to companies effectively disregarding consumer privacy rights. The scathing revelation forced Amazon to attempt an IOAIL certification for its popular vacuum cleaner. Unfortunately, the device was rejected certification as Amazon were unable to satisfy basic requirements due to third-party stakeholders refusing to reveal in entirety the extent of data usage, collection, and processing. Though the product is still available on the market, sales of it have been affected as consumers now aware of the situation, have moved to alternative products such as the *Xiaomi Mi Vacuum*, which although a similar product also sports an IOAIL 2 Class mark. From this it can be viewed that, if anything, the presence of this certification provides users a sense of security, even if that security falls

within the alleged 'grey areas' of the IOAIL system. Giving the consumer the choice and knowledge of a technology's functionality and noted pitfalls.

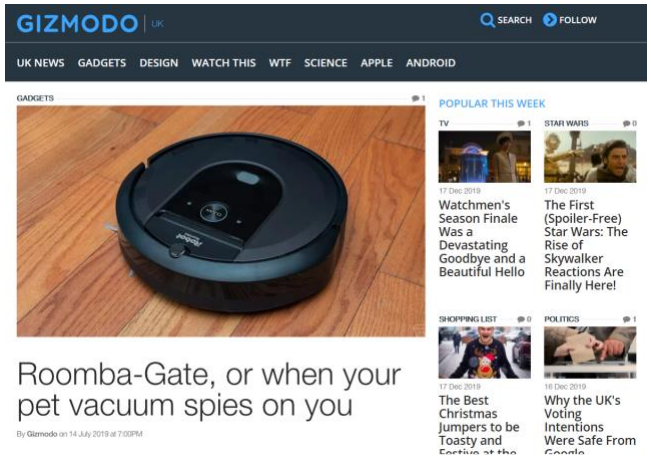

Figure 5: Ripples of the IRobot Roomba controversy broke the internet, many articles like this one shown explained the sourcing of dimensional data of many people's homes thereby affecting the product's sales.

The IOAIL Procedure

The IOAIL requires a company or applicant to apply though a certification procedure supplying evidence in a report and access to the AI systems. As previously mentioned, certification can be granted even with limited disclosure and access, rewarding those that take part in the certification that companies will publicly acknowledge that some form of AI is present in the product. AI Security Ops, the suppliers of the security system in place at Market Square are not IOAIL certified making public announcements that due to the secure nature of their security packages access is restricted to in-house only, for security reasons and to maintain commercial secrecy, further exacerbating the grey areas and the lack of transparency of these AI systems. Working with the council's permission our access to the system was limited as a third-party, though we were able to work with the additional information the council had as consumers of the security package. Avoiding legal action from AI Security Ops, we applied for certification through the council, as owners of the security package they had legal rights to attempt to certify and provide visual marks for a product in a public space. However, we suspected as with many AI technological companies the ownership of technology is often a complex case of ownership between client and company, especially when it's an ongoing cloud-based service, thereby we suspected that we would not be able to acquire knowledge or a mark of the full attributes of the AI system in Market Square.

That said, applying for certification was done over the course of 2 weeks whereby an IOAIL inspection team attempted to assess the technology through the provided data which included redacted documentation and limited links to sources provided by AI Security Ops. It was understood that a legal team had perused the documents beforehand to assure any stakeholders associated with AI Security Ops were cleared. Still the IOAIL assured us that even though the information provided was limited it was enough to begin assessment for an IOAIL mark. The following is a brief of the resulting report from IOAIL.

Report

A spectrum analysis was provided which assessed 5 core aspects of any AI's ranged ability (**Figure 6**). These attributes include the location of data processing

in regards to the AI and/or product, the scope of processing as a static or trainable AI, data provenance as open or restricted, the various types of data collection, and intrinsic labour done by the AI, a concept established out of the Rights of Users of AI Act for the transparency of AI activity in relation to users.

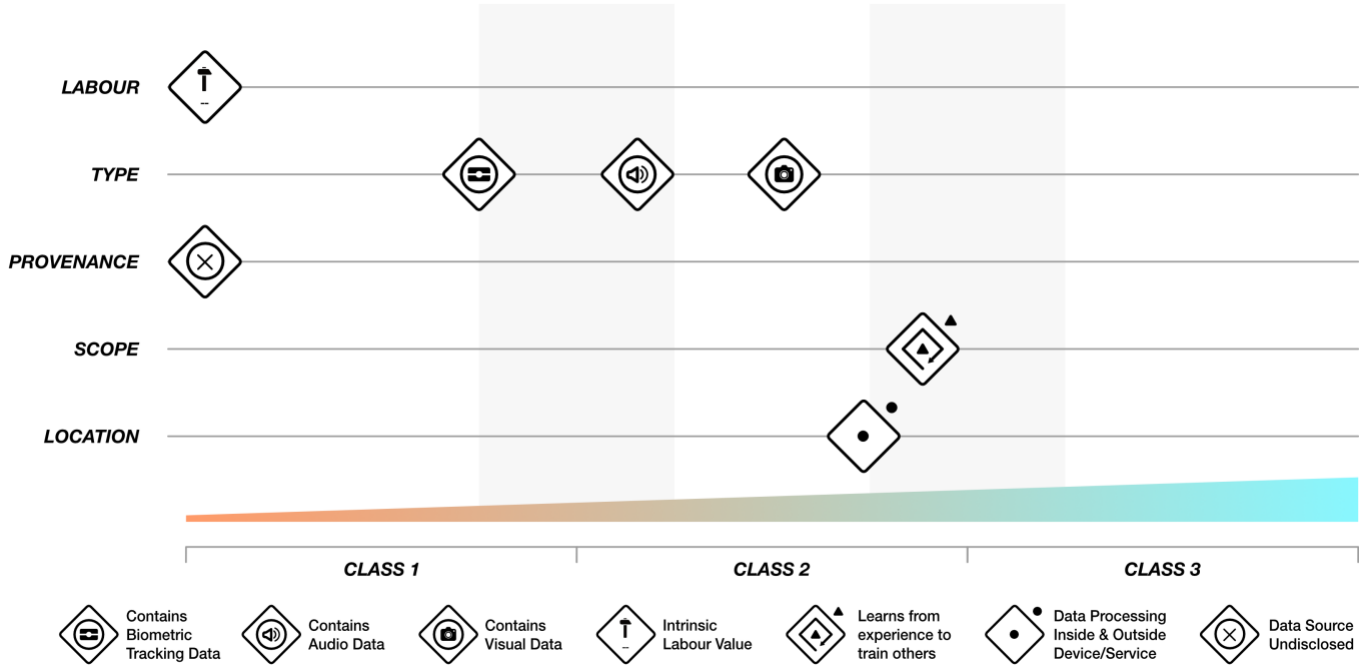

Figure 6: Spectrum Analysis of the security system in place at market square. Each icon represents an attribute of the AI technology and is positioned on the spectrum dependant on the disclosure of particulars. The iconography is a standardised system produced by IOAIL and is of an abstract nature reminiscent to those found on clothes labels for laundry instructions.

Overall, the spectrum analysis showed that the AI present in the security system as far as could be assessed by the information provided, used on-board and cloud-based sources for processing. Though, they

both were unreachable as the provenance showed it to be proprietary technology not open to public and any meddling with the physical cameras would have been unlawful, the IOAIL could only assess these attributes through alternate sources linked to the technology which the inspection team were able to find. Included into this is the fact that the AI present in the cameras is, as far as being advertised, able to learn from experience and train it's cloud-based processing. This further presents the AI as, what the IOAIL class as, Second Adaptive Scope of AI processing. This means that not only are the cameras capable of assessing the data it is presented via on-board and cloud-based technology, it also is able to train their algorithms to understand the data better through machine learning. Albeit, this should be taken with a grain of salt as it is unknown as to what form of processing is being done behind the scenes consequently permitting claims of it to be 'better'. Moving on, the spectrum also presents the 3 forms of data the AI collects in ascending order of transparency provided: Tracking, Audio, and Visual data. Finally, there was no information present that could help the IOAIL asses the intrinsic labour achieved by the security system which though an abstract concept still requires hard data in order to be assessable.

Four data points were definitively in grey-areas in this spectrum: the two data types Tracking and Audio, as well as the full scope and location of data processing. This stems from insufficient information regarding the processing capabilities of the technology used by AI Security Ops in their security systems. According to IOAIL regulations a system requires a certain score accumulated through the assessment in order to be given a class mark upon the spectrum. In its current state given the information provided to the IOAIL, the spectrum analysis presented the security system with a IOAIL 2 Class mark, but we were told the score it accumulated was mostly acquired through grey-area attributes. In other words without further information in resolving grey-areas the classification could be considered inconclusive and the system would need to mention the presence of these grey-areas. The most common method seen of going around this problem is by adding 'fine-print' information alongside any IOAIL certifications.

Findings and Discussions

The above information once acquired was presented to Lancaster City Council who are currently in discussion of whether to put up the IOAIL suggested signage with mention of grey-areas as public knowledge or not. Their stance is that where on the one hand this situation presents a means for making AI enhanced spaces more legible in a democratic manner, the particulars of this case could raise public suspicion further around these ambiguous grey-areas of data usage within technology. Nonetheless, the council presented assurances of looking into the matter further and perhaps entering discussions with AI Security Ops regarding the legibility of their technology further.

Curious to know how the public would react to this, brief interviews were conducted on Market Square with willing participants. The reaction was mixed with most favouring the need for legibility and an interest in knowing where and how this data was used. Though we were assured that the data was used for security purposes, the full extent was not provided in the report therefore it cannot be fully asserted. That said, there

were also people less interested in the matter saying if it kept them safe then it was probably good.

In all, this research was intended to shed light on the matter of legibility of AI systems. From the manner in which the public has reacted to the presence of enhances IoT capabilities particularly with ever growing advances in technology, it can be said that the legibility of AI is in effect a pressing issue. Where certifications methods such as those provided by IOAIL do exist, how efficient this system of standardising IoT legibility truly is can be presented for debate. The manner in which legibility is assessed could still be considered incomplete with no means of attaining information from third-parties the process of certification is hindered. The end result is a compromise of sorts with the standardising body using 'grey-areas' as a way to circumnavigate the bureaucracy of tech. In order for any manner of clear and transparent AI legibility to be attained it would be necessary to firmly establish its importance for not just the public but also third-party stakeholders who can essentially put a wrench in the workings so to speak.

Conclusions

This research paper and the associated artefacts presented here is a design fiction[6]. This design fiction is a continuation of researching AI legibility through design and the development of AI iconography through a Research-through-Design (RtD) methodology. The research intends to create visual communication accessible within the context of HCI to enhance AI legibility and to define what approaches to be wary of. The artefacts and iconography in this paper are, by no means, intended to solve or conclude the challenge of making AI legible and transparent, but as a speculative exercise in the adoption of a system of iconography to communicate the, at present, opaque operations and parameters of AI infused products. The world built in this paper concentrated on several aspects that may become apparent if AI iconography and the application of marks existed such as; how the process would be established in an official capacity of a certification body and the process of ratification, the relationship between client and company that still own rights to cloud-based services and the objective of keeping trade secrets, market forces that underpin adoption of enhanced AI legibility, to the simple necessity of having transparent technology, especially in public areas, that collect vast amounts of data on people that ultimately go on to govern people's lives. Future research is concerned with creating a more in depth and tangible AI iconography system that communicates effectively AI inner functionalities by reaching out to purveyors of AI services and policy makers. We also intend to create further design fictions to consider future measures to be taken responding to the accelerating reality of an AI engulfed, but hopefully, transparent world.

Acknowledgements

We thank the Lancaster City Council for fictional access to the operational AI security package located in Market Square and taking part in our research.

This research has been supported through the Ucanny AI project funded through the PETRAS National Centre for Excellence in Cybersecurity EPSRC project EP/S035362/1.

References

[1] Bryson, J., & Winfield, A. (2017). Standardizing ethical design for artificial intelligence and

autonomous systems. *Computer, 50*(5), pp. 116-119.

[2] Burell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society.* https://doi.org/10.1177/2053951715622512

[3] Coulton, P., Lindley, J., Akmal, H, & Pilling, F. (2025). The mysterious case of AI Iconography, CHI 2025.

[4] Koutamanis, A., Halin, G., & Kvan, T. (2007, April). Information standardization from a design perspective. In *CAADRIA 2007,* (pp. 1-8).

[5] Liker, J. K., Collins, P. D., & Hull, F. M. (1999). Flexibility and standardization: test of a contingency model of product design–manufacturing integration. *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, *16*(3), pp.248-267.

[6] Lindley, J. & Coulton, P. (2016). *Pushing the limits of design fiction: the case for fictional research papers.* Paper presented at CHI 2016, pp.4032-4043.

[7] Lindley, J., Coulton, P., Akmal, A, H. & Pilling, F. (2019). *Researching AI Legibility through Design*. Paper to be presented at CHI 2020.