Making AI Infused Products and Services more Legible.

Franziska Pilling PhD Candidate, Lancaster University, f.pilling@lancaster.ac.uk, ORID: 0000-0001-5324-0575.
Haider Ali Akmal Research Associate, Lancaster University, h.a.akmal@lancaster.ac.uk, ORID: 0000-0001-9578-3578
Joseph Lindley Lecturer in Design Research, Lancaster University, j.lindley@lancaster.ac.uk, ORID: 0000-0002-5527-3028
Adrian Gradinar Lecturer in Smart Home Futures, Lancaster University, a.gradinar@lancaster.ac.uk ORID: 0000-0001-5416-6207
Paul Coulton Professor of Speculative Design, Lancaster University, p.coulton@lancaster.ac.uk, ORID: 0000-0001-5938-4393

Abstract
The increasing availability of large data sets has initiated a resurgence in Artificial Intelligence (AI) research after a period of reduced funding caused an AI winter in the late 1980s. Today AI is integrated into a wide variety of so-called smart products and services to personalise user experiences. Smart Technologies are typically designed for ease of use, with their underlying complex procedures purposely obfuscated, leading to misconceptions about how AI works from a lack of legibility. Using a Research through Design approach, the authors address the challenge of AI legibility through iconography to enhance agency and understanding of AI and data-infused interactions.

A Process of Obfuscation; The Challenge
AI and data collection have become core activities within the cloud-based services empowering smart thermostats, streaming services, and AI assistants, such as Alexa, and are becoming increasingly ubiquitous in our daily activities. However, the underlying operations relating to AI and data in these networked products are predominantly illegible such as the user's voice data being recorded to train AI assistants in use. This illegibility diminishes user agency and their ability to negotiate the transactional nature of such activities, as obtaining functionality is often conditional on agreeing to provide data [1]. Designs that intentionally obfuscate operations relating to AI and data do so for a variety of reasons: under the aim of being Human-centred, by simplifying the interaction to avoid overloading the user with auxiliary information [2], or for institutional protection, to conceal intellectual property, and

in some cases deceptively [3] to collect data without explicit consent [4]. While the Cambridge Analytica scandal made some Facebook users more aware that data produced by their interactions with Facebook is recorded, it is less obvious how AI is processing this data and for what purpose by Facebook and the third parties they supply it to. This lack of legibility presents many challenges, resulting in users' perception of AI being more heavily influenced by science-fiction renderings of Artificial General Intelligence as sentient 'thinking' machines like The Terminator (1984), rather than the mundane reality of narrow AI using machine learning utilised by services such as Netflix to predict viewing preferences. This dichotomy is the 'Definitional Dualism of AI' [5] and highlights part of the challenge of making the functional elements of AI more legible to users. While the notion of legibility is promoted in many frameworks to encourage 'better' strategies for AI implementation [6], they do not provide any examples of how this is achieved in practice which is the focus of this article.

Researching AI Legibility through Design

To address the challenge of how AI legibility might be achieved in practice, we adopted a Research through Design (RtD) approach to develop a set of AI icons to communicate AI operations, diffuse the complexity, and raise user awareness of AI functions and how data is being employed in the products they use.

As a practice-based approach, RtD provides a generative aptitude to 'explore, speculate, particularise, and diversify' the challenge in hand towards manifesting findings into rich research artefacts [7]. To this end, our research interweaved methods from disciplinary-diverse perspectives, for example: Human-Computer Interaction for theories for richer inter-relationships between users and computers [8]; Semiotics - modelling the constructs of signifiers [9]; to amalgamating design and AI research, for instance, how design can implement trust in AI products [10] given the bias issues prevalent in training data [11]; and promoting legibility and explainability of AI functions over transparency which is often more related to making AI products auditable by specialists [12].

Rather than focussing on text-based descriptions of AI, which tend to use technical jargon or require expert knowledge, we assessed the current levels of AI legibility by surveying AI iconography [13] as the visualisation of information is often used to increase accessibility. This showed that, although some AI imagery tries to represent the underlying system such as neural network (Fig. 1a) or others highlighted its use, such as face detection (Fig. 1b), the vast majority play into AI's Definitional Dualism by showing human-like machines (Fig. 1c & d) exacerbating misconstrued perceptions of human intelligence existing in AI. The survey also highlighted that current AI imagery rarely communicates the intricacies of how an AI functions and in what context, emphasising the need to develop a new visual approach to enhance AI legibility.

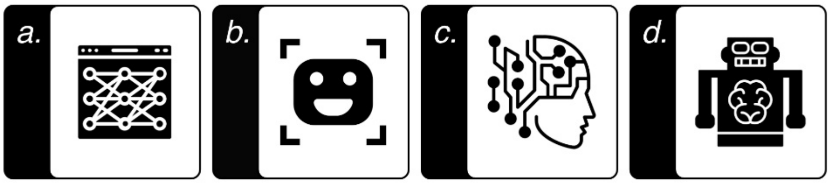

Figure 1. Examples of current AI iconography.

Our initial design response resulted in seventeen graphical icons (Fig. 2), which individually and in their grouping detail a particular operational function or feature we identified as crucial rudimentary components to communicate. These were: *Learning Scope* communicating how the AI adapts is a fundamental factor for human-AI interaction [14]; *Data Provenance* to declare the source of the training data, as data reflects the AI and therefore its trustworthiness [15]; *Processing Location* to define where processing is occurring, which impacts users' perception of accountability [16] and *Training Data Type*, which is a granular account of the type of data used to train the AI to reduce opacity, bias and increase trust. The final Icon is *Intrinsic Labour,* a critical reflection of the monetisation of data through the commodification of users and their interactions. These icons can illuminate an AI's operational activity without engaging with the specifics of the AI's implementation, such as using neural nets.

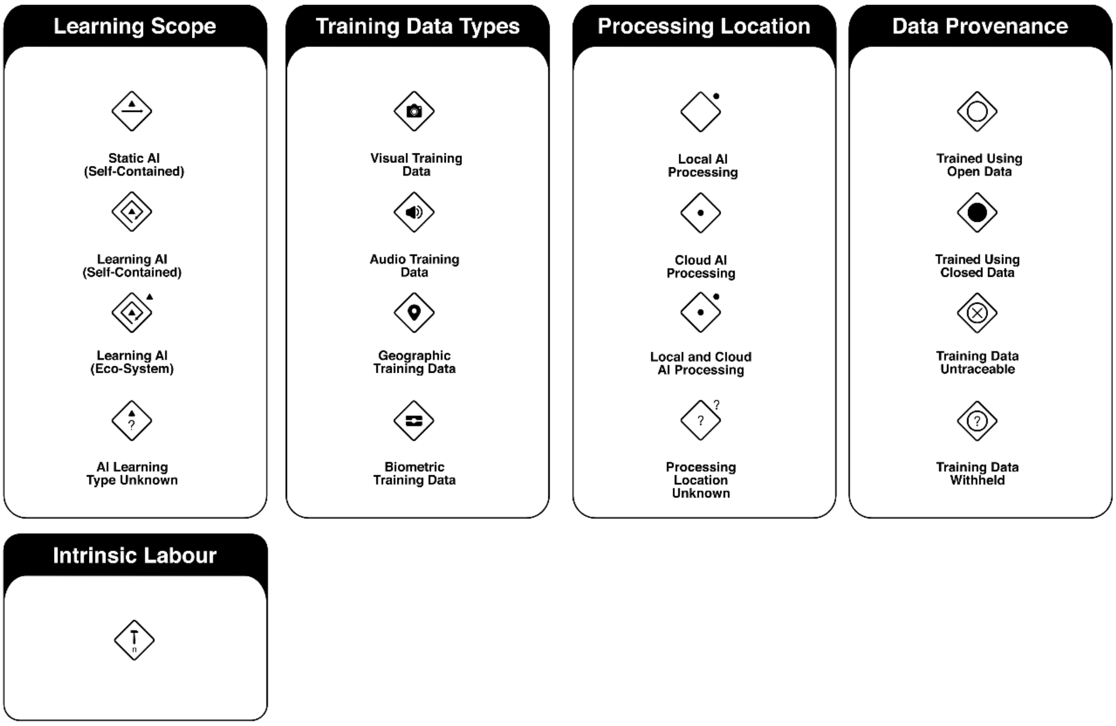

Figure 2. The first iteration of icons.

These icons can then be used in contextualised combinations to map and communicate an AI's 'ontological constituent' articulating an AI's implementation to the user [17]. We designed three different icon styles during our ideation, pictorial, textual, and abstract variants. Our initial evaluation settled on the abstract icons as they best avoided the pitfalls previously discussed.

Further, these abstract icons conformed to the principles of semiotics, where icons often hybridise symbolic, indexical, and iconic categories together to communicate the intended concept. These icons were designed with a diamond shape to retain uniformity between them and be easily combined in different configurations. Note, the abstract AI icons were also inspired by laundry care labels initially created in 1971 because while we may not always take notice of these abstract icons, or indeed always understand their meaning, they provide a means of making legible how we can most easily maintain a working relationship with our clothes.

An icon works if the user can match the interpretant to the intended object – or, in the instance of a digital thing, which is arguably more challenging to capture having no 'conventional representation' [18] – its concept or the implication [19]. However, through semiotics, a plausible representation can emerge [20] and can be embedded within an icon to become 'iconic' [21] and be utilised by would-be users. To ascertain how well the icons performed, we engaged in iterative and interactive evaluation workshops with potential stakeholders, including end-users, academics and industry practitioners who deploy AI.

Testing Intuitiveness and Icon Development

Due to Covid, the workshops were conducted online using a bespoke set of playful activities produced using the game engine Godot. A playful approach allowed us to ease participants of all knowledge levels towards discussing potentially complex ideas outside their experience. The initial workshops evaluated the icons, with forty-seven participants' and in the following paragraphs, we note our observations and how these influenced the creation of the second set of icons.

The first exercise of the workshop was matching the icons to their descriptors. It quickly became apparent that the *Training Data Types* and *Processing Location* icons were the most intuitive, with the highest correct matches. The success of the *Training Data* icons was because they contained well-known 'iconic' signifiers, such as an audio speaker for audio training data, and 'symbolic' signifiers [22] that have become embedded into our society, such as the geographic pin for geographic data.

As all seventeen icons were present in one setting, participants developed non-verbal reasoning tactics to match icons and their textual descriptions and then use the resemblance to one another to collate into the corresponding groupings. For the *Processing Location* icons, participants commented on the fact that they were able to decipher these individually and group them as the icons employed the symbolic element of a circle to represent processing and were specifically positioned either inside or out of the aforementioned 'AI diamond' to represent internal or cloud processing.

From the results of the workshop discussions, we identified the problematic aspects of some icons. For example, to communicate an AI trained once offline, the icon Static AI was presented with a triangle used to symbolise 'learning' and beneath a directional arrow pointing to the right. Many participants observed that the arrow suggested movement rather than stasis. This arrow was changed to a triangle enclosed by a diamond shape that sat inside the icon's AI diamond for the second icon iteration. This configuration better conformed to the group's symbology where an open arrow path in a diamond shape symbolised continuous learning; hence a closed diamond accentuated static (Fig. 3).

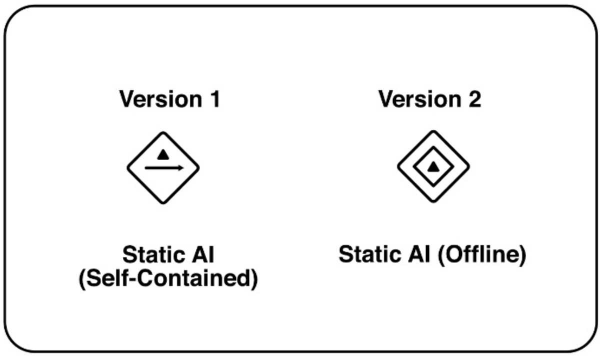

Figure 3. Comparison of Static AI icons.

Framing AI Concepts

The workshops also tested how well the AI's relationship to data was being communicated and how we framed these concepts. We attempted to design the icons to be expansive enough yet not ambiguous. However, this proved problematic in the way we sought to frame the concepts drawn from discussions amongst those working in AI research. Consequently, the *Data Provenance* category was redefined as *Training Data Origin* as this proved more understandable during workshop discussions. Additionally, the training definitions we initially framed were found to be vague and beyond the scope of knowledge for everyday users because of the niche and specialist terminology we used. For instance, the concept *Trained Using Open Data* where participants often asked what open data meant, to which facilitators would answer "data to be audited by an external body, to determine whether the data is representative of the activity it is being applied to." Therefore, in the second iteration, we reframed what specific icons were communicating to be more accessible for general users. In particular, 'open data' changed to *Trained Using Auditable Data,* and most importantly, we created an icon for *Trained using User Data* which participants were often most concerned about.

The workshops and the icons as tools brought about supplementary observations regarding how users developed a better understanding of AI operations, as many participants remarked that they had more knowledge and critical awareness of AI technology and data gathering techniques from completing the workshop. This comprehension led to a handful of participants claiming they would immediately disable the location permissions on their devices. We also identified which icons and their concepts needed to be self-explanatory. For example, the concept *Learning AI (Eco-system)* was confusing and required an understanding of what the ecosystem metaphor meant in the context of AI learning; thus, we framed the concept *AI to AI Learning*.

Developing New AI concepts

During the first workshops, it became apparent that the AI's overall application was not communicated, with participants seeking this information rather than how it was doing it. Consequently, the category of *AI-Assisted Decisions* was designed for the second iteration, expanding the number of icons to Twenty-One (Fig 4.). This icon set proved problematic to design an abstract pattern due to existing understandings associated with particular terms; for example, using a crystal ball to signify prediction would play into the saturated discourse on technology and magic [23], which we wanted to avoid. The icons we settled with uses letters, although these are not ideal for translation into other languages.

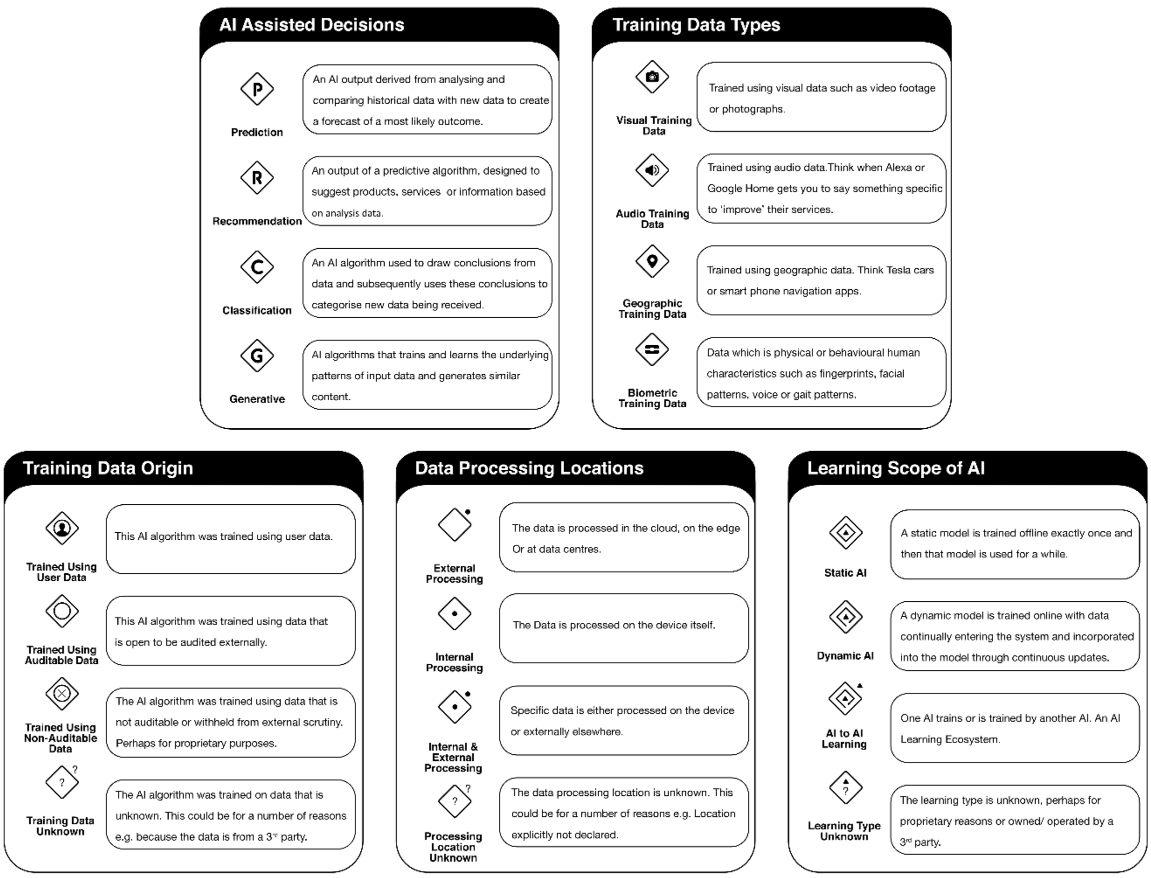

Figure 4. Version two of AI icons (minus Intrinsic Labour).

Designing a User Priority Arrangement

Discussing with participants the information they felt was most relevant to them resulted in speculating a hierarchical order for the icons. First, noting the 'presence of AI' with a subsequent breakdown of the AI in question and be in an order that conformed to users' 'priority of information' (Fig .5). In this way, users could quickly access the most relevant information, and for users wanting a technical understanding, this would be detailed further down the hierarchy. Thus, a user can decide how much information they would need to make a conscious decision of how to interact with an AI device. Together, the icons abate the indeterminacy of interaction, and the hierarchy uniformly organises information in a way that further diminishes ambiguity but at a comfortable depth of detail for the user. In the second set of the workshops, we asked participants to rate the icons based on information they wanted to know about their device to establish an order of importance. Interestingly, the *Training Data Types* as a category ranked the most important, with participants particularly wanting to understand what type of personal data an AI was recording and learning from. Correspondingly, *Trained Using User Data* was the common highest choice, whereas the remaining *Training Data Origin* category and *AI Learning Type* were considered less important to know and be placed further down the hierarchy.

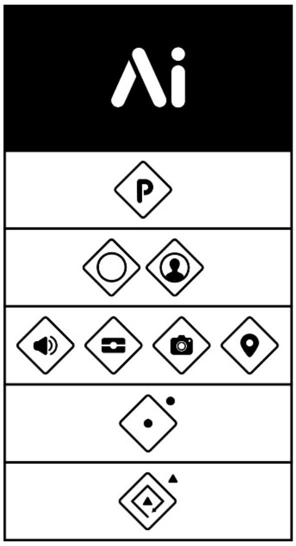

Figure 5. An example of the AI Hierarchy System.

Second Iteration Performance

The second iteration of icons had a high success rate of being intuitively matched to the textual descriptors, with several participants matching all correctly. Granted, the participants had both the concept and the icon to work with, rather than arbitrarily guessing what the icon could mean. However, the empirical testing we chose to conduct reflects the reality of launching icons out into the 'wild,' where the semiotic design of the icons establishes a visual vocabulary for an immaterial concept. Over time these icons would become concretised in meaning and familiarised due to their tested and considered design.

The icons that proved to be less intuitive based on forty-five participants results were the more abstract icons of *AI Learning Type* and *Training Data Origin*. Nevertheless, participants who did match these icons commented that crucial parts of the icons could be identified and considered a visual representation. For instance, the rotational arrows in *AI Learning types* resembled "a round of training or an iteration of learning".

It must be recognised, while the icons can facilitate a greater understanding and legibility of AI operations, for many systems, it is a dubious claim to fully explain how a particular decision was reached as AI computing is un-interpretable. AI processes attempt to determine indeterminacy, whereas indeterminacy is internal to the process itself amid also attempting to bring the indeterminacy of the world and lived experience into computation [24]; however, what is known can be made legible and communicated.

Unaccounted Concepts

With experience in completing the workshops and the rapid advancement of AI research, we have begun designing accompanying icons that mirror current ethical and responsible considerations (Fig 6). The icon *Intrinsic Labour* offered the opportunity to critically discuss and question the larger impacts of using AI technology, which is too often far removed from the immediate interactions and outputs a user experiences when using AI-assisted devices. Icons can address the notion of responsibility, enabling us to make better choices and empower responsible technology.

Figure 6. Test examples for responsible AI icons.

The first developmental icon is *Human in-the-Loop*, signifying that some form of human labour occurred for the AI to operate. This icon could acknowledge the oft misconception that an AI operation is purely performed through computation. The reality is that a human's hand, expertise and sensitivity is still an integral part of the automation revolution with activities such as data labelling [25]. 'Human Out-of-the-Loop' systems can evolve beyond human intelligibility and perform tasks in a way that is literal, although ultimately wrong. Frank Lantz's paperclip game shines a fantastical light on this, whereby an AI destroys the world, making paperclips as requested, illuminating that AI and human values deviate [26].

An icon of this prominence would aid in developing more critical awareness and scope to fortify workers' protection within these technological sectors through policy and educate users about AIs' accurate foundations, guiding users to opt for more coherent and rational AIs.

While the second concept is still in its preliminary stage, we are keen to create an icon, or set, that promotes the ethical use of intimate data that has potential benefits for the common good, such as patient data to prevent or predict disease. Ethically approved AI applications or services could be certified with these *Cooperative Icons* to communicate to users that their data is used for good on neutral terms, which would embody trust in interacting with these services.

Conclusion

The research presented here is not expected to solve the evolving challenge of legibility but to demonstrate the potential of design-led responses and illustrate our RtD methodology of interweaving interdisciplinary perspectives to design graphical AI icons to communicate AI's intangible and complex functions for more informed use. We outlined the developing iterations of the icons to demonstrate the multifaceted challenge of legible AI. This is achieved by steering away from AI's indeterminacy nature – in that rarely can we say for certain why an AI has reached a particular decision or by relying on and broadcasting convoluted expert knowledge. The design challenge for legible AI is to create an accessible and uniformly constructive AI lexicon to demystify AI and its potential implications on a user.

Acknowledgements

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. Mortier R, Haddadi H, Henderson T, McAuley D, Crowcroft J. Human-Data Interaction: The Human Face of the Data-Driven Society. 2015.
2. Norman D. The Invisible Computer: Why Good Products Can Fail, the Personal Computer is So Complex, and Information Appliances are the Solution. MIT; 1998.
3. Burrell J. How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data Soc. 2016.
4. Zuboff S. The Age of Surveillance Capitalism. London: Profile Books Ltd; 2019.
5. Lindley J, Akmal HA, Pilling F, Coulton P. Researching AI Legibility through Design. CHI 20 Proc 2020 CHI Conf Hum Factors Comput Syst. 2020:13.
6. Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. SSRN Electron J. 2020.
7. Gaver W. What should we expect from research through design? In: Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12. Texas, USA: ACM Press; 2012.
8. Bowers J, Rodden T. Exploding the interface: experiences of a CSCW network. In: Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93. Amsterdam, The Netherlands: ACM Press; 1993.

9. Peirce CS. On a New List of Categories. In: Hoopes J, editor. Peirce on Signs. University of North Carolina Press; 1991. p. 23–33.
10. Arnold M, Bellamy RKE, Hind M, Houde S, Mehta S, Mojsilovic A, et al. FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity. 2019.
11. Angwin, J, Larson J, Mattu S, Kirchner L. Machine Bias There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica; 2016.
12. Lindley J, Coulton P. AHRC Challenges of the Future: AI & Data. 2020.
13. Lindley et al. [5]
14. Amershi S, Weld D, Vorvoreanu M, Fourney A, Nushi B, Collisson P, et al. Guidelines for Human-AI Interaction. 2019;13.
15. Arnold et al. [11]
16. Rader E, Cotter K, Cho J. Explanations as Mechanisms for Supporting Algorithmic Transparency. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems [Internet]. Montreal QC Canada: ACM; 2018.
17. Pilling F, Lindley J, Akmal HA, Coulton P. Design (Non) Fiction: Deconstructing/Reconstructing The Definitional Dualism of AI. Int J Film Media Arts. 2021;6(1):6–32
18. B Ferreira J, Noble J, Biddle R. A case for Iconic Icons. 2006.
19. Barr P, Noble J, Biddle R. Icon R Icons: User interface icons, metaphor and metonymy. New Zealand: Victoria University of Wellington School of Mathematical and Computing Sciences; 2002.
20. Ibid.
21. Ferreira et al [18].
22. Ibid.
23. Davis E. Techgnosis: Myth, Magic, Mysticism in the Age of Information [Internet]. Harmony Books; 1998
24. Fazi MB. Contingent Computation: Abstraction, Experience, and Indeterminacy in Computation Aesthetics. London: Rowman & Littlefield; 2018.
25. Natarajan S, Mishra K, Mohamed S, Taylor A. Just and Equitable Data Labelling towards a responsible AI supply chain. aapti institute; 2021.

26. Rogers A. The Way the World Ends: Not with a Bang But a Paperclip. Wired. 2017 [accessed 08/07/2021]; Available from: https://www.wired.com/story/the-way-the-world-ends-not-with-a-bang-but-a-paperclip/