Cutting together-apart entangled conceptualisations of (machine) learning and ethics: assemblage of a 'bricolage-pentimento' artefact through an ethico-onto-epistem-ological approach

Michael Wolfindale
Student number: 1684487
Year: 2023
MSc in Digital Education, University of Edinburgh

Table of Contents

Word count: *16497 words excluding appendices and references.*

i. List of Tables & Figures

Figures

Tables

ii. Declaration of own work

Having reviewed the programme handbook and course-specific guidance on good academic practice, and reflected on the nature of plagiarism and the ethical representation of academic knowledge, I confirm that this piece of work is submitted without any misappropriation of sources.

iii. Abstract

Artificial intelligence (AI) degree programmes commonly include machine learning courses covering techniques shaped by specific notions of learning. The ethics of such technologies are covered as part of many AI programmes, although are less often integrated into core machine learning courses. For this research, I first explore how students studying machine learning conceptualise 'learning' in general, and any correlations with machine learning techniques. Secondly, I consider potential ethical implications of these conceptualisations of learning in the context of AI education and development of machine learning technologies. Interrogating this entanglement of notions of (machine) learning, educational practice and ethics—involving human, nonhuman, material and abstract (more-than-human) entities—is challenging if one engages honestly with its complexity. Furthermore, I argue that integration of critical discussion of ethics into AI education design and practice requires an experimental combination of approaches—built upon a methodology emphasising difference and complexity over sameness and simplicity. Taking as a point of departure Barad's 'ethico-onto-epistem-ology'—acknowledging that knowing, being and ethics are intertwined—and considering new insights that multimodal social-semiotics can bring, I employ multimodal ethnographic arts-based approaches while drawing upon new materialist theories. Such anti-anthropocentric theories—which recognise more-than-human entanglements within an unstable and evolving 'research-assemblage'—may appear opposed to human-centric multimodal social-semiotics. However, I draw upon Haraway's diffractive methodology to read multiple theoretical insights *through* one another, spreading meaning in unpredictable yet productive emergences. Thus, multimodal artefacts exploring conceptualisations of learning were produced through participatory conversations with machine learning students. Then, employing arts-based techniques including 'bricolage' and 'pentimento', a new artefact was assembled through iterative diffractive analysis and (re-)configurations of existing materials. Drawing on Barad's 'agential cuts', this 'bricolage-pentimento' artefact visualises a 'cut' of continually changing flows of affective relations between more-than-human components through which students' conceptualisations of learning emerge, highlighting computational, mathematical and behaviourist themes. Ultimately, multiple insights into conceptualisations of (machine) learning and potential ethical implications were gained, whilst articulating how multimodal and new materialist perspectives might be brought together productively through a diffractive methodology. Such insights and methodologies may inform future transdisciplinary AI programme design, deeply integrating critical (multimodal) consideration of theories of (machine) learning—and their social, historical and political contexts—with discussions of ethics.

iv. Acknowledgements

I am eternally grateful to the student participants who freely offered their time, the academic staff who helped me to invite students to participate, and technical staff who helped me to set up the necessary infrastructure. My time on the University of Edinburgh's Digital Education programme has been an experience I will treasure, and I am especially indebted to my supervisor Dr Ben Williamson, as well as all the staff and students of the programme who have been so crucial in guiding me through. I write with special thanks to my Mum and Dad, and continue to be inspired by my brother and sister—who remain in our hearts though they are sadly no longer with us. Particular thanks go to Anna, David, and Robin Purnell who have and continue to be forever supportive. Finally, this thesis is dedicated to Kandida, without whom it would not have been possible, and to our wonderful cats Bertie, Bernie, and Biffo for their moral support throughout.

1. Introduction

As 'artificial intelligence' (AI) technologies continue to permeate and shape daily life, there is a growing interest in the involvement of AI systems with education (Chen et al. 2020a; Chen et al. 2020b; Hwang et al. 2020; Laupichler et al. 2022). Debating the ethics of such systems is increasingly part of many AI and data science programmes, which often also include computer-science focused courses in machine learning (Saltz et al 2019). Although there have been attempts to integrate ethics into computer science curriculums (Grosz et al. 2019; Cernadas and Fernández-Delgado 2021), ethics is most commonly studied separately from machine learning courses, such as in stand-alone 'AI ethics' courses often taught from a philosophical or legal perspective (Lim et. al. 2023; Weerts and Pechenizkiy 2022). It is less common for ethics to be deeply integrated into core technical machine learning courses (Saltz et al. 2019).

The field of machine learning typically involves techniques and models shaped by specific notions of 'learning'. One technique is 'reinforcement learning', which draws on radical behaviourist theories (Knox et al. 2020), and is derived from areas of psychology concerned with animal learning (Sutton and Barto 2018). Traditional approaches focused on the 'biomimicry' of human brains (Dicks 2016; Floreano and Mattiussi 2008), while more recently modelling biological structures more widely through, for example, artificial neural networks (Nwadiugwu 2020). It is unclear to what extent the social, historical and political context of these ubiquitous approaches—and underlying contested ideas of learning—are framed by AI educational programmes. Neither is the impact upon students' own conceptualisations of learning apparent, nor any subsequent effect on how students of machine learning may decide what constitutes a 'ethical' AI system—which they may later be involved with developing.

Thus, there is a need for critical reflection on 'ethics' to be more commonly and deeply *integrated* into AI educational programmes and core machine learning courses. To achieve this, a diverse range of methods and educational activities is required that goes beyond the limitations of self-contained AI ethics courses (Lim et al. 2023). If one takes the stance that ethics should be 'infused' through core machine learning courses (Saltz et al. 2019: 21), then transdisciplinary activities should interrogate conceptualisations of learning which inform machine learning. Furthermore, investigation is needed into how the framing of specific notions of 'learning' in machine learning courses affect students' conceptualisations of 'learning' and 'ethics' in the context of AI systems.

Thus, for this research project, I set out to explore the questions in Table 1.a:

How do students studying machine learning conceptualise 'learning'?
What are the potential ethical implications of these conceptualisations of learning in the context of AI education and the development of machine learning technologies?

Table 1.a: Research questions.

These questions could be addressed from many perspectives, depending on one's theoretical standpoint. However, if one acknowledges the complexity of an entanglement of human, nonhuman, material and abstract (more-than-human) entities involved in these conceptualisations of learning, and their intersection with ethics, this has a profound impact on the underlying theoretical and methodological frameworks employed. Chapter 2 reviews several frameworks relevant to addressing these questions, from differing theoretical perspectives. Given the unique insights multimodality can offer for the representation of knowledge (Bayne et al. 2020), and the importance of being open to modes beyond text while exploring broad concepts such as 'learning' (Kress 2011), the production of multimodal artefacts is employed in the research design. The approach of multimodal social-semiotics (Kress 2009) is discussed in chapter 2, due to its relevance for analysing the meaning made through these signs/artefacts.

However, the anthropocentric tendency of multimodal social-semiotics to centre on the agency of the human individual (Newfield 2018) presents limitations if one intends to explore complex relations between more-than-human entities entangled in notions of (machine) learning, educational practice and ethics. Such exploration is a key aim, and thus several relevant new materialist concepts—taking anti-anthropocentric approaches to explore agency and ethics in the context of such relations—are also reviewed in chapter 2. This begins by introducing several underlying theoretical concepts—the 'assemblage' and 'affect' (Deleuze and Guattari 1988)—and contextualising how they inform new materialist sociological frameworks such as the 'research-assemblage' (Fox and Alldred 2015; 2016), comprising of flows of affective relations between more-than-human components. Then, theories relevant for exploring ethical implications in the context of more-than-human assemblages—Karen Barad's (2007) agential realism, intra-action and ethico-onto-epistem-ology and N. Katherine Hayles' (2017; 2022) cognitive assemblage framework—are covered.

While these different perspectives may each offer productive insights into exploring conceptualisations of learning, tensions are also revealed. Such tensions centre around differences in how the (human) subject, agency and representationalism are conceptualised, and where the primary focus for enquiry should lie—ranging from the human-centric focus of multimodal social-semiotics to the anti-anthropocentric approach of Barad's (2007) agential realism and intra-action. However, drawing upon Denise Newfield's (2018) bringing together of multimodality and posthumanism through a 'diffractive' approach, the potential of a diffractive methodology (Haraway 1997)—to read differing (even seemingly opposed) insights *through* one another—is discussed next. Such an approach emphasises difference over similarity (Jackson and Mazzei 2012; Mazzei 2014), and aims for the affirmative production of new insights, patterns and questions (Barad 2014).

Building on the diffractive approach introduced in chapter 2, the rationale for employing a diffractive methodology for this project is covered in chapter 3. Here, I set out my theoretical point

of departure—drawing on Barad's (2007) ethico-onto-epistem-ology—and how this aligns with my approach. A diffractive methodology brings together a mix of theories—in this case, multimodal social-semiotics and new materialist thinking—and this informs the mix of methods employed for data collection and analysis. Furthermore, my approach is laid out in context of two broader project goals. The first goal is to explore my research questions from multiple theoretical perspectives—primarily multimodal social-semiotics and new materialisms—and gain insights into students' conceptualisations of learning and potential ethical implications. Producing generalisable data is not a primary goal—this is limited by the sample size but, moreover, the second goal is to articulate how multimodal approaches might be productively brought together with new materialist thinking through a diffractive methodology. Given the experimental nature of this methodology which deliberately engages with complexity, I draw upon danah boyd's (2008) ethnographic approach, suggesting researchers remain open for new questions to be revealed by data as the research progresses. This aligns with a diffractive methodology and approach to analysis, which are 'rhizomatic' and may unpredictably lead in many directions in a non-linear fashion (Mazzei 2014)—as discussed further in chapter 3.

While my starting point is the framework of multimodal social-semiotics, concerned with the representation of knowledge through sign-making, gaining deeper insights into the wider social context calls for an ethnographic approach to data collection; indeed, as Gunther Kress (2011) argues, these different approaches can be complementary (Kress 2011). As such, I draw in particular upon multimodal ethnographic approaches (Flewitt 2011; Kress 2011; Pink 2011; 2015; 2021) to inform my methods for data collection. Thus, in-depth multimodal (online) conversations with students of machine learning were employed—these collaborative participatory conversations, which go beyond scripted interviews, allowed space for deeper exploration of conceptualisations of learning and ethics through multiple modes beyond only text; multimodal artefacts were produced during the conversations. Such conversations took place over the course of several months, while students were studying courses in machine learning, and field notes were produced during this time. These conversations, multimodal artefacts and field notes were subject to multiple stages of analysis, initially from a multimodal social-semiotic perspective, and then subsequently a 'diffractive' analysis. This stage of analysis involved reading the initial analysis and data fragments alongside/through new materialist theories—theories which acknowledge a complex entanglement of more-than-human components—in order to gain fresh insights and reveal new questions.

Drawing on arts-based techniques, including 'bricolage'—creating a new artwork from existing materials—and 'pentimento'—layered compositions from different authors and times (OED 2022), these analysis stages culminated in the assemblage of a new artefact. This 'bricolage-pentimento' artefact was formed through combining visual (re-)configurations of existing artefacts produced at different times, annotated with new insights such as the visualisation of affective relations connecting components, evoking Barad's (2007) concept of 'spacetimemattering' where

entanglements of space, time and matter are intra-actively 'cut together-apart' through ongoing iterative (re-)configurations. This draws upon Barad's (2013: 18) notion of the 'agential cut', whereby 'subject' and 'object' are entangled but momentarily separated through a 'cutting together-apart', making analysis of phenomena (and presentation of the bricolage-pentimento artefact) methodologically possible. By visualising this 'cut' of affective flows between more-than-human components through which students' conceptualisations of learning emerge, this multimodal artefact supports my findings and makes visible insights gained from the diffraction of data through new materialist ideas. The theoretical underpinnings of this approach are detailed in chapter 3, while findings from each stage of analysis are presented together with the 'bricolage-pentimento' artefact in chapter 4.

Ultimately, while the project and methodology was experimental, unpredictable and rhizomatic in nature, two aims were achieved. First, insights were gained from both multimodal social-semiotic and new materialist perspectives into conceptualisations of (machine) learning and the potential ethical implications. Second, it is methodologically articulated how a diffractive approach can productively bring together multimodality and new materialisms to explore and reveal important questions and considerations. Such insights, methodologies and demonstration of multimodal arts-based approaches may contribute to informing the design of future AI programmes and educational activities—making visible the social, historical and political contexts surrounding theories of (machine) learning, and deeply integrating these considerations with discussions of ethics through transdisciplinary approaches. Bringing these strands of thought together, the closing chapter 5 summarises the conclusions gained from this project, and suggests how they might inform such transdisciplinary educational design.

2. Literature Review

Having outlined the aims of the project in chapter 1, and briefly pointed to the potential insights that multimodal approaches may bring when exploring conceptualisations of learning, this chapter begins with a review of relevant multimodal approaches which inform both the research design (chapter 3) and analysis (chapter 4) employed. However, multimodal approaches can tend to be human-centric (Newfield 2018) and, given a key aim is to explore relations between human, nonhuman, material and abstract entities, the following sections review several anti-anthropocentric new materialist theories relevant for exploring this complex more-than-human entanglement.

Some of these new materialisms derive their theoretical frameworks from the concepts of Gilles Deleuze and Félix Guattari (1988), such as notions of *assemblages* and *affect*. These concepts are first introduced, and then contextualised with their relation to new materialist theories particularly relevant to addressing the objectives of this project. These include the new materialist sociologist framework developed by Nick J. Fox and Pam Alldred (2015; 2016), drawing on the notion of affective flows within the *research-assemblage*—an unstable more-than-human entanglement of researcher, object of research, apparatus, technologies and theories (Fox and Alldred 2015: 404). This informs my theoretical point of departure (see chapter 3), as does Barad's (2007) *ethico-onto-epistem-ology*, *agential realism* and concept of *intra-action*, covered next. Finally, Hayles' (2017) *cognitive assemblage* framework—drawing upon Baradian and DeleuzoGuattarian theory—is reviewed. Each framework offers differing interpretations of new materialist theory, yet each offers a productive lens through which to analyse more-than-human relations and their ethical implications.

Various theoretical commonalities and differences are revealed in this chapter, and tensions are particularly apparent between the seemingly opposed viewpoints of human-centric multimodal approaches and anti-anthropocentric new materialisms. However, this chapter concludes by introducing how a diffractive methodology (Haraway 1997) offers possibilities for reconciling such differences through constructive and deconstructive analysis (Barad 2014), gaining valuable insights from multiple theories. A diffractive approach provides a means to engage productively with this multiplicity of ideas, revealing new questions while acknowledging some inevitable ambiguity (Mazzei 2014), and feeds into the rationale for this project's methodology, detailed further in chapter 3.

2.1. Multimodal approaches

If 'many modes matter in representing academic knowledge' (Bayne et al. 2020: 47), methods of analysis that incorporate multimodality may be important to consider when researching conceptualisations of (machine) learning. This is particularly relevant if students should choose to

convey their understanding of concepts through modes other than text—for example, drawings or images. An openness to this variety of modes, Kress (2011: 246) argues, may in fact be important for broader questions such as 'what is learning?'.

While multimodality is a vast domain encompassing many theoretical frameworks (Pirini et al. 2018), I will focus on the multimodal social-semiotic approach set out by Kress (2009), which draws upon Michael Halliday's (1978; 1984) perspective on semiotics. Its particular relevance for this project lies with its focus on analysis of semiotic resources in a social context—in this case, (multimodal) representations of conceptualisations of learning produced through conversations with students.

Kress's (2009) multimodal social-semiotics is primarily concerned with meaning arising in social environments through sign-making. These signs—fusing meaning and form—can exist in any mode, where a mode is a 'socially shaped and culturally given semiotic resource for making meaning' (ibid.: 79); images, moving images, writing, speech and music are examples of modes used in communication and representation. The *sign* is a representation—always a *partial* representation of a concept, phenomenon or object (the *signified*), but a *full* representation of the interests of the sign-maker at that moment in time, having been translated into 'apt means of representing it' (the signifier) (ibid.: 71). The interests of the maker of the sign—which, in the case of this research project, refers to both the participants conveying their conceptualisation of learning, and myself as researcher conveying my analysis—materialise as metaphors. Thus, if one takes a social-semiotic approach to communication and representation, 'all signs are metaphors, always newly made' (ibid.: 55).

In contrast with many of the new materialist approaches discussed later, multimodal social-semiotics focuses on the (human) individual, their socially shaped histories and their agency (Kress 2009). One such example is explained by Kress (2011), where a science teacher asks students a broad question about cells and the student responds verbally (one sign) and then via a drawing (another sign). Modes are not limited only to speech, writing or drawings, however; observing gestures and the production of three-dimensional objects may also be relevant when taking a multimodal social-semiotic approach (ibid.). Furthermore, the students' agency in this sign-making process would be 'seen as not arbitrary but as *motivated* by principles of selecting an apt form for the desired meaning' (Kress 2011: 247, emphasis in original).

A multimodal social-semiotic approach, then, might offer valuable insights for analysis of conceptualisations of learning represented by students. Furthermore, Kress (2011) goes on to argue that, to understand wider social characteristics, an ethnographic approach may complement a multimodal social-semiotic account; it is covered further in chapter 3 how multimodal ethnographic (Pink 2011; 2015; 2021) approaches inform my methodology. How, though, can one address the tendency for multimodal social-semiotic approaches to be human-centred—with a focus on the individual and their agency in the act of sign- and meaning-making (Newfield 2018;

Kress 2011)? Anti-anthropocentric new materialist theories may offer alternative lenses through which to analyse students' conceptualisations of learning, but how might these apparently incompatible approaches be reconciled? The following sections introduce the relevance of various new materialist frameworks for exploring these conceptualisations while acknowledging complex more-than-human relations, and then show how different apparently conflicting approaches might be reconciled through a diffractive approach.

2.2. DeleuzoGuattarian materialisms and the (research-)assemblage

It is a key intention of this project to interrogate conceptualisations of learning (and potential ethical implications) beyond human-centred representationalist perspectives, and explore their emergence through entanglements of human, nonhuman, material and abstract entities. The anthropocentric limitation of multimodal social-semiotics, therefore, highlights a need to shift our focus to the *relationality* of such more-than-human entities. Furthermore, the tendency of humanist ethics to position ethics within universalising discourses can confuse issues of accountability/responsibility (Taylor 2018), as discussed further in section 4.2.3. Valuable anti-anthropocentric insights into such matters are offered by the new materialisms discussed below, which often draw upon DeleuzoGuattrian materialism (Fox and Alldred 2016), focusing on the relationality of more-than-human entities through the notion of the assemblage.

Deleuze and Guattari (1988: 21) build on the metaphor of the nonhierarchical and nonlinear 'rhizome' (shown in Figure 2.a), which 'connects any point to any other point', and 'has neither beginning nor end, but always a middle', to conceptualise the assemblage. Moving beyond subject-object dualisms, the assemblage questions divisions between fields of reality, representation and subjectivity, and instead 'establishes connections between certain multiplicities drawn from each of these orders' (ibid.: 22-23). Replacing the more traditional concept of agency, and drawing on the notion of affect derived from Spinoza, assemblages are held together by the capacity for assembled relations to 'affect or be affected' (Deleuze and Guattari 1988: xvii). Thus, the focus shifts to the (affective) *flow*s between (more-than-human) assemblages.

The DeleuzoGuattrian focus on *relations between* entities, rather than individual entities themselves, informs both the methodology for this project (in chapter 3) and the new materialist theories employed as a framework for analysis (presented in chapter 4), such as Hayles' (2017) cognitive assemblages (discussed below). However, as Barad and Hayles are keen to point out, their theories are not wedded to DeleuzoGuattrian paradigms; rather they offer differing theoretical interpretations and valuable lenses through which to analyse these relations, and such difference is handled productively through the diffractive methodology introduced below.

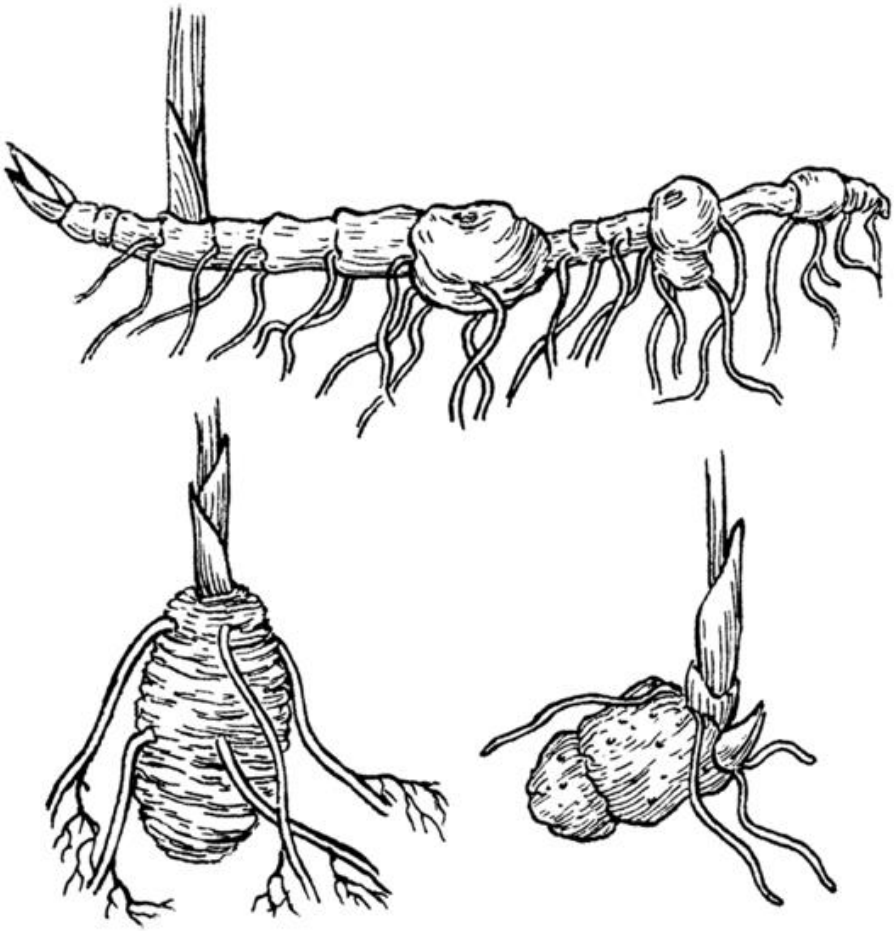

Figure 2.a: The rhizome.
Image by Pearson Scott Foresman (public domain).

These ideas also form the basis of Fox and Alldred's (2015; 2016) new materialist sociologist framework, whereby affective flows are continually branching in a rhizomatic fashion, within a constantly (re-)developing and always becoming *research-assemblage*, an intertwining of researcher, object of research, research apparatus, technologies and theoretical frameworks. This in part informs the theoretical point of departure for this project, as does Barad's (2007) ethico-onto-epistem-ology which is introduced next.

2.3. Barad's agential realism, intra-action and ethico-onto-epistem-ology

Barad's (2007) new materialisms offer a productive framework through which to address the aforementioned need to move beyond human-centred representationalist perspectives, and her notion of *intra-action* informs Hayles' (2017) cognitive assemblage framework (covered below).

However, while both have commonalities with DeleuzoGuattarian materialisms in their focus on the relations between more-than-human entities, they diverge in their view of representationalism—while Hayles does not altogether deny its benefits, Barad seeks to move beyond it through a performative approach, as discussed below.

Barad (1996) draws on quantum mechanist Niels Bohr's challenging of Newtonian/Cartesian observer independence and differentiation between object/observer (Fox and Alldred 2016), to shift the focus of enquiry from individual entities to *phenomena*. Through her concept of agential realism, Barad (2007: 140, emphasis added) argues that individual entities are not pre-existing nor do they *inter*act; rather, phenomena are 'produced through complex agential *intra*-actions of multiple material-discursive practices'.

This ontological shift has profound implications, too, for Barad's (2007: 178) conceptualisation of agency. Building on Judith Butler's (1990) theory of performativity, Barad (2007) moves beyond representationalism to argue how agential realism reconceptualises material and discursive practices in a *performative* relationship. For Barad (2007: 49), performativity questions the ontological separation of representations and the entities being represented; rather, the focus for enquiry is the nature of 'the practices or performances of representing' and 'the productive effects of those practices and the conditions for their efficacy'. Thus, in the context of this project, and in contrast with human-centric multimodal approaches, the focus for enquiry is shifted away from the agency of research participants towards these performative practices.

Rather than someone or something having agency, then, what matters is 'the enactment of iterative changes to particular practices'. In every moment, there exist 'changing possibilities for (intra)-acting', and this results in an 'ethical obligation to intra-act responsibly in the world's becoming, to contest and rework what matters and what is excluded from mattering.' Thus, the focus is the 'boundary-drawing practices' through which 'the differential boundaries between humans and nonhumans, culture and nature, science and the social, are constituted' (ibid.: 140). These repeated (re-)drawing of boundaries are known as 'apparatuses of bodily production' (Haraway 1988: 595), in contrast with the notion of the observing apparatus. Barad (2007) terms these repeated performative boundary-making practices *agential cut*s which, through intra-active (re-)configurations, enacts momentary *agential separability* between 'subject' and 'object' within phenomena.

This reconceptualisation of agency has profound implications for how I methodologically explore and analyse the emergence of students' conceptualisations of learning and the surrounding ethical implications (see chapter 3). The notion of 'cutting together-apart' (Barad 2013: 18)—whereby entities are not independent, but through continual intra-active (re-)configurations enact agential separability within phenomena—is in fact what makes it methodologically possible to present a 'cut' of my project findings (see chapters 3 and 4).

Furthermore, my theoretical point of departure for this project is built upon Barad's (2007: 185) *ethico-onto-epistem-ology*—which rejects the separation of ontology and epistemology and argues that 'we don't obtain knowledge by standing outside the world; we know because we are of the world…part of the world in its differential becoming'. Since each changing possibility for each intra-action matters, and we are part of this constantly becoming world, we are co-implicated along with other human, nonhuman and material bodies (Taylor 2018). Thus, Barad's (2007) ethico-onto-epistem-ology appreciates that knowing and being are intertwined with ethics and, for this project, taking this entanglement with ethics forward impresses the need to interrogate the phenomena of students' conceptualisations of learning and the entangled intra-actions with more-than-human entities through which they emerged. Moreover, as researcher I am mutually implicated in this by presenting a 'cut' of findings for this project. This starting point informs my methodology—elaborated in chapter 3.

2.4. Hayles' cognitive assemblage framework

Hayles (2017; 2022) notion of cognitive assemblages provides a framework relevant for analysis of the more-than-human entanglement through which students' conceptualisations of learning emerge, and a lens through which to consider ethical implications. Key to Hayles' (2017: 22) contribution is a focus on cognition—'a process that interprets information within contexts that connect it with meaning'. Drawing on the DeleuzoGuattarian assemblage, cognitive assemblages are constantly mutating arrangements of biological and technological systems and actors through which this information flows, 'effecting transformations through the interpretive activities of cognizers operating upon the flows' (ibid. 118).

Hayles' (2017: 1) cognitive assemblage framework lays out a rethinking of cognition which recognises the existence and power of 'nonconscious cognitive processes', opening up the possibility for *noncognizers* to possess agential powers. Hayles (2017: 31-32) differentiates between *cognizers*—actors which are 'embedded in cognitive assemblages with moral and ethical implications'—and noncognizers—material forces and objects with a lack of capacity to make choices, performing instead as *agents*. This category of agents also includes those objects acting as 'cognitive supports', or that can 'perform cognitive tasks when suitable constraints are introduced'.

Diverging from DeleuzoGuattarian paradigms and drawing on Barad's (2007) theory of intra-action, Hayles (2017: 75) seeks a middle ground between 'Deleuzian becomings and cognition, subjectivity and higher consciousness', between 'material processes and modes of awareness', whereby 'intraactions connect sensory input from the internal and external environments ("events") with the emergence of the subject ("entities")'. In common with Barad's (2007: 141) rejection of individual "things"—events or entities—as the primary ontological focus, Hayles (2017: 75-76) argues that our focus for enquiry should instead be 'the points of intraaction, the dynamic

and continuing interplays between material processes and the structured, organized patterns characteristic of consciousness.'

Seeking a bridge between pro- and anti-Deleuzian perspectives, Hayles (2017: 75) argues that denying representation entirely may underplay 'possibilities for nonconscious cognition and representational actions'. Rather than position the Deleuzian view of contingent assemblages between entities—intensities and forces preceding the individual, and from which everything else originates—as contradictory to the view that the individual subject is a 'pre-existing entity upon which forces operate', Hayles (2017: 75) in fact interprets these two perspectives as two sides of a DeleuzoGuattarian assemblage, each with their own insights.

Drawing on Rosi Braidotti's (2013) conceptualisation of the subject, Hayles (2017: 76-77) seeks a mediating approach between 'forces/intensities and subjects/organisms', whereby 'nonconscious cognition is the link connecting material forces to us as subjects…without requiring that subjects be altogether erased or ignored as agents capable of political actions'. This balancing act between approaches—where the subject and representationalism are not altogether denied—diverges somewhat from Barad's (2007) move beyond representationalism towards performativity, where subject and object are contingent, do not preexist, and emerge only through intra-activity.

What, then, do cognitive assemblages—fluctuating human/nonhuman/computational collectivities through which information, interpretations, and meanings circulate (Hayles 2019; 2022)—mean when considering ethics? Building on the argument that, in order to be labelled as an ethical actor, an entity must possess the attributes of interactivity, autonomy and adaptability (Floridi and Sanders 2004), Hayles (2022: 4) transforms these attributes into a 'logical hierarchy'. The capacity for *interactivity* between entities—such that 'ethical effects' are created through an interactive system acting 'with intentions that may affect others'—sits at the foundation of this hierarchy since, 'without it an entity, like a hammer or knife for instance, would simply be inert'. Adding to this is *autonomy*—'the ability to make self-directed choices and thus to become morally accountable for them', and at the top is *adaptability*—which contributes 'the capacity to change based on feedback from the environment and from previous choices, opening the possibility of learning through trial and error and, ultimately, to evolving into new states of becoming.'

According to Hayles (2022: 2), agency is distributed among entities, at multiple decision-making levels, with different level of responsibility and accountability depending on the capacity of the actors 'to interact, be autonomous, and adapt to changing circumstances'. Rather than pretend there is a simple way in which we can conceptualise agency and control in this context, we must instead acknowledge and address the inherent complexities. This ethical framework for cognitive assemblages provides a particularly productive framework for analysis, and is taken forward in chapter 4.

While numerous differences have been highlighted between multimodal approaches, and the new materialist frameworks of Barad and Hayles, the following sections introduce how a diffractive approach might affirmatively reconcile such differences.

2.5. A diffractive approach

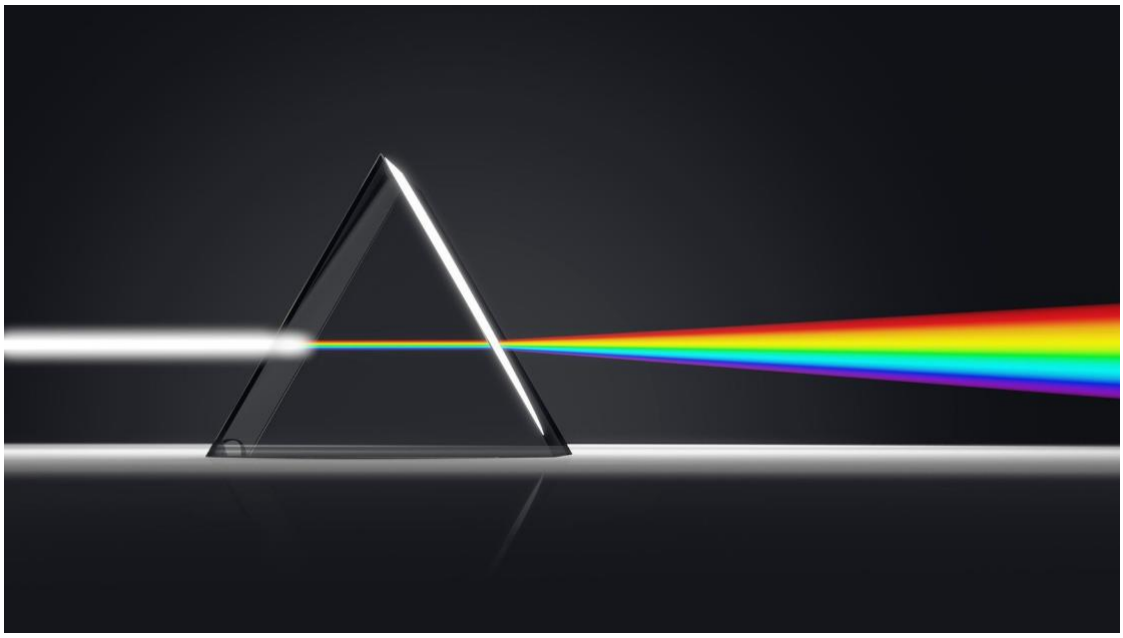

Figure 2.b: Diffraction of light through a prism.
***Image:** Daniel Roberts, Pixabay.*

Diffraction, a term from physics, refers to the phenomenon of new patterns emerging when different (light, sound or other) waves combine and experience obstructions (Barad 2007), as shown in Figure 2.b and Figure 2.c. Haraway (1997) offers diffraction as an alternative metaphor to reflexivity; rather than mirroring or reflecting, diffracting produces new patterns from the interference of different waves. By emphasising difference, a diffractive methodology destabilises 'the tropes of liberal humanist identity work necessary in conventional qualitative research: the subject, interpretation, categorical similarity' (Mazzei 2014: 743). Crucially, for the purposes of this project, it offers a productive way in which to combine different theories which may appear incompatible—namely the aforementioned human-centric multimodal and anti-anthropocentric new materialisms—through diffractive analysis.

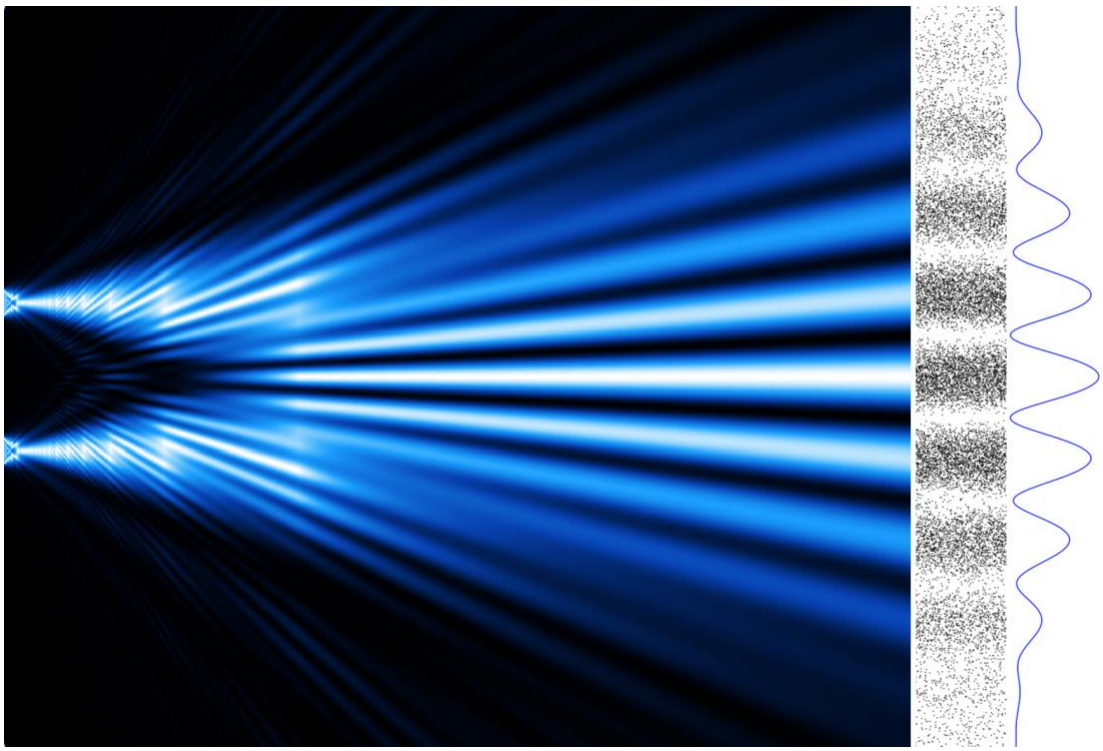

Figure 2.c: Two-slit diffraction.
***Image:** Alexandre Gondran. **Licence:** CC BY-SA 4.0.*

Lisa A. Mazzei (2014) lays out a framework for diffractive analysis which offers a means to engage with multiple theoretical insights and moves beyond a conventional search for categorical similarity through coding. In the same way that 'breaking apart a ray of light shows up its many components and indeed co-constructs these components in the process of diffraction' (Hickey-Moody and Willcox 2019: 3), a diffractive analysis breaks open the data and threads it through theory (Jackson and Mazzei 2012). Such insights may 'interfere' with one another producing 'new patterns of understanding-becoming' (Barad 2014: 187), similarly to how waves constructively and destructively interfere (see Figure 2.d). However, diverging from the notion of destructive interference (where waves are "cancelled" out), Barad (2007: 205, emphasis added) calls instead for a 'de*con*structive' analysis which emphasises 'accountability for the particular exclusions that are enacted' and our mutual 'responsibility to perpetually contest and rework the boundaries' intra-activity co-constituted between humans/nonhumans.

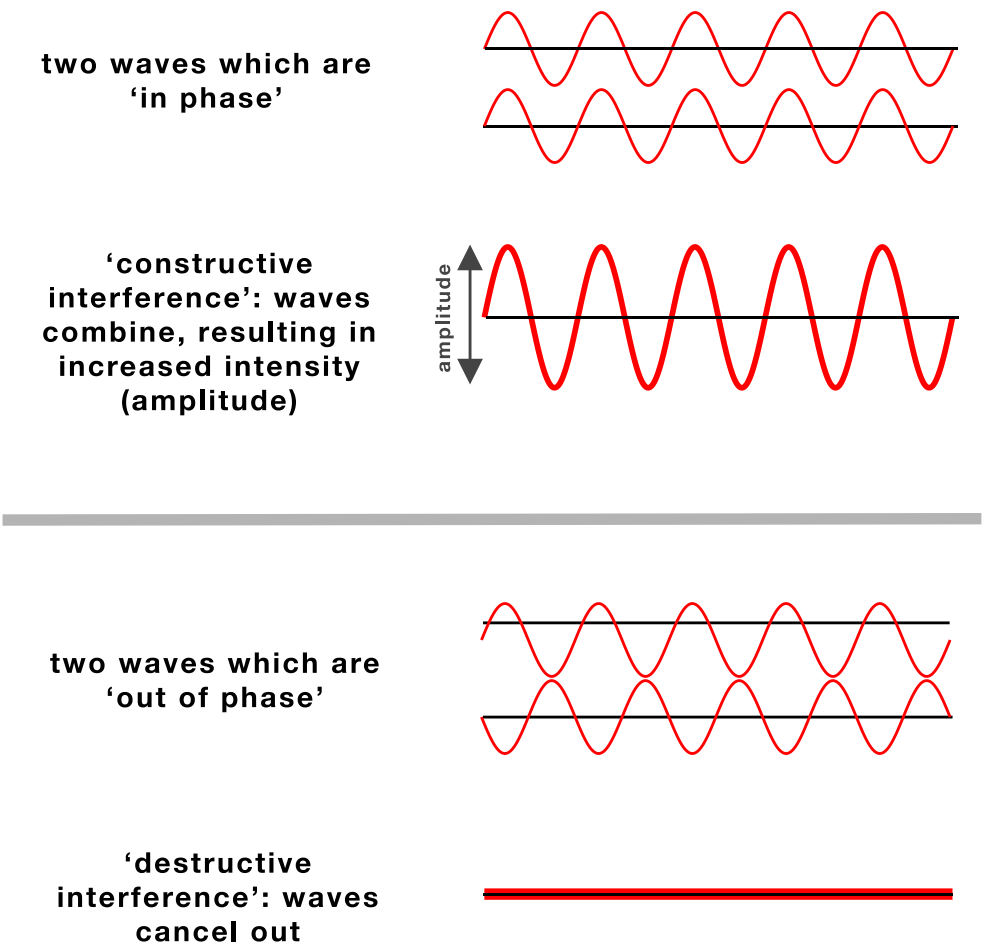

Figure 2.d: Constructive and destructive interference.

Adapted from image by Haade, Wjh31 and Quibik (Wikimedia Commons). Licence: CC BY-SA 3.0.

Returning to the context of this project, multimodal social-semiotics offers unique insights when exploring representations of students’ conceptualisations of (machine) learning through different modes, while the new materialisms of Barad (2007) and Hayles (2017) offer differing but productive ways to consider the emergence of such conceptualisations through the more-than-human entanglement of the research-assemblage. Furthermore, Hayles’ (2022) cognitive assemblage framework provides a way to engage with the ethical complexity of such an entanglement. A diffractive analysis brings these strands together by reading different theoretical insights through each other (Barad 2007), engaging productively with tensions (Kuby and Christ 2018). By taking a ‘rhizomatic form’ that may lead in new and unpredictable directions, new insights, connections and questions are generated (Mazzei 2014: 743). Thus, these multiple theoretical insights provide a framework for considering students’ conceptualisations of learning, and the surrounding ethical implications, while a diffractive approach brings these together and informs my methodology, covered next.

3. Research Design & Methodology

Having reviewed literature relevant to addressing the questions in Table 1.a, this chapter demonstrates how the theories covered there inform my research design and methodology. My theoretical point of departure builds upon Barad's (2007) ethico-onto-epistem-ology, introduced in chapter 2, and this is covered first. Building upon this theoretical framework, and drawing on arts-based techniques, this leads into laying out the methods employed in this project. Finally, the processes followed for participant recruitment, data generation and analysis are outlined in detail.

3.1. Methodological approach

3.1.1. Towards an ethico-onto-epistem-ological approach

Building on the concepts introduced in chapter 2, I take as a point of departure the unstable and evolving 'research-assemblage' (Fox and Alldred 2015), where the 'object of research' and 'researcher' are entangled. This calls into question the notion of a single objective 'reality' or 'true' knowledge that is 'out there' to be ascertained. Furthermore, I take a 'nomadic' approach to enquiry where we are continually in a 'process of becoming' (Braidotti 2011: 41), an approach that acknowledges my position as researcher while simultaneously *centring* myself—my biases, power and influences—and *decentring* myself by being attentive to the influence of other agents—human, nonhuman and material—in the research-assemblage (Sidebottom 2019).

Considering the implications decentring the human has for conceptualising (machine) learning, I think in terms of 'cognitive assemblages'—entanglements of humans, nonhumans and computational media (Hayles 2017). Drawing upon on Barad's (2007) notion of 'intra-action', I reconsider the concept of 'interaction' (implying that independent entities exist and interact); instead, 'it is through specific agential intra-actions that the boundaries and properties of the "components" of phenomena become determinate and that particular embodied concepts become meaningful' (Barad 2003: 815). 'Phenomena' do not refer to 'epistemological inseparability of observer and observed, or the results of measurements'—rather, they are '*the ontological inseparability/entanglement of intra-acting "agencies"*' (Barad 2007: 139, emphasis in original).

Thus, my primary ontological focus is phenomena 'produced through complex agential intra-actions of multiple material-discursive practices or apparatuses of bodily production' (Barad 2007: 140). These apparatuses give definition to specific concepts over others through boundary-making agential cuts. Interrogation of phenomena for this project is only possible through such cuts, which are performative in nature (Barad 2013). As researcher, I am not an impartial observer; rather, I am entangled in the setting up of such apparatuses—neither innocent nor fully 'in control' (Barad 1996; Völker 2019).

Taking this as a point of departure, however, makes problematic dualisms inherent in humanist qualitative research (Lather and Pierre 2013)—human/nonhuman, mind/body, mind/matter, nature/culture, subject/object, ontology/epistemology. Thus, I take forward Barad's (2007) 'ethico-onto-epistem-ological' approach (introduced in chapter 2), which acknowledges that knowing, being and ethics are intertwined and where 'ethical concerns are not simply supplemental to the practice of science but an integral part of it…*values are integral to the nature of knowing and being*' (Barad 2007: 37, emphasis in original). This not only adds extra weight to the argument that 'ethics' should be integral to core machine learning courses (Saltz et al. 2019), but has profound implications for my research design and methodology, as discussed below.

Returning to my research questions (see Table 1.a), and taking forward the theories introduced in chapter 2, this ethico-onto-epistem-ological approach insists upon an interrogation of the cognitive assemblages of humans, nonhumans, material and abstract entities through which conceptualisations of learning emerge. The becoming of these conceptualisations is entangled with social, scientific, political and ethical practices (Barad 2003), and data-driven machine learning technologies and concepts. These cognitive assemblages do significant work in our everyday lives; yet, ethical frameworks that honestly account for these are still relatively few and far between (Hayles 2022).

Thus, my broad goals are twofold. Firstly, I look to interrogate this entanglement and gain productive insights into conceptualisations of learning within the research-assemblage. Second, I take a reflexive view of my own research within this assemblage, in order to gain wider methodological insights into how one might productively engage with these complex issues from (apparently conflicting) multimodal and new materialist perspectives. Thus, my methods must acknowledge the 'agentic assemblage of diverse elements that are constantly intra-acting, never stable, never the same' (Lather and Pierre 2013: 630), and these are introduced in the following section.

3.1.2. Methods: cutting together-apart entangled conceptuallsations

If one takes an 'ethico-onto-epistem-ological' approach, then, what are the implications for studying the 'entangled web of scientific, social, ethical, and political practices' (Barad 2003: 813)—where knowing, being and ethics are intertwined? According to agential realism (see section 2.3), the main ontological unit is not separate pre-existing objects (with fixed boundaries and properties) that interact; instead, it is the phenomena 'produced through complex agential intra-actions of multiple material-discursive practices or apparatuses of bodily production' (ibid.: 139-140). These apparatuses are 'boundary-drawing practices' through which 'the differential boundaries between humans and nonhumans, culture and nature, science and the social, are constituted' (Barad 2007: 140). In contrast with the 'Cartesian cut'—which distinguishes between 'subject' and 'object'—

these ‘apparatuses enact agential cuts that produce determinate boundaries and properties of “entities” within phenomena.

Considering that ‘art, photography, collage and other affective means of responding can provide new insights while also bringing in embodied elements’ (Sidebottom 2019), my research design is inspired by arts-based and multimodal approaches that draw upon new materialist thinking (see Albin-Clark et al. 2021; Carlyle 2020; Lasczik Cutcher 2018; Hickey-Moody 2016; Hickey-Moody and Willcox 2019; Marsh 2017; Warfield 2016)—recognising entangled ‘more-than-human’ entities within the research-assemblage. Thus, in order to explore machine learning students’ conceptualisations of ‘learning’, and drawing on ideas from multimodal ethnography (Pink 2011; 2015; 2021), I set up online asynchronous conversations with students at machine learning courses running at a university where I worked as an educational advisor. As well as text chat, the conversations involved the production of multimodal ‘artefacts’ based on their conceptualisations of (machine) learning and questions of ethics (see Figure 3.d). The artefacts—generally drawings or images—along with conversation transcripts, were first subject to a multimodal social-semiotic analysis. Field notes were generated throughout, and artefacts, transcripts and field notes were finally ‘diffracted’ through new materialist theories. These stages build on the concepts introduced in chapter 2, and the process is detailed below.

While human-centric multimodal approaches may seem incompatible with anti-anthropocentric new materialist thinking, I bring together insights gained from different approaches through a diffractive methodology (see section 2.5). While such insights may ‘interfere’ with each other, they are read through one another (Mazzei 2014) in order to generate new insights and questions. As detailed below, this methodology informs not only analysis of student artefacts, conversation transcripts and field notes, but also the creation of a new ‘bricolage-pentimento’ artefact.

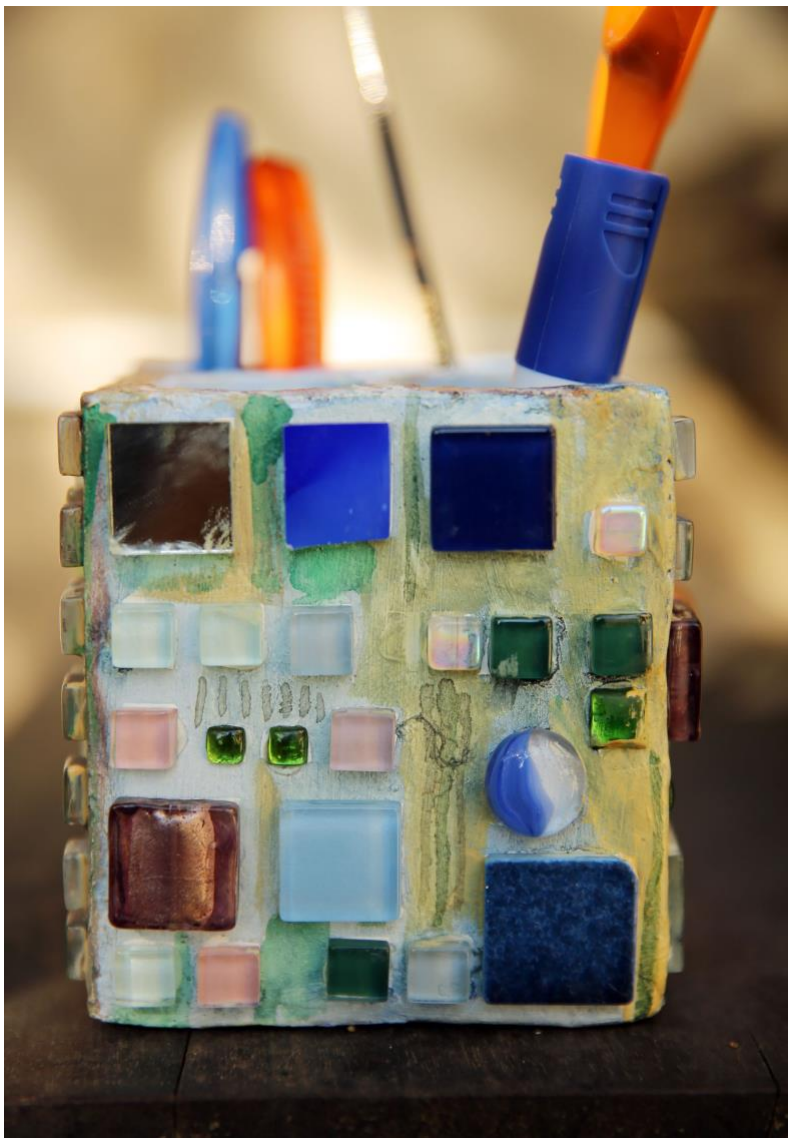

Figure 3.a: ***Bricolage d'été***.
***Isaszas*. *Licence:* *CC0 1.0*.**

Using the technique of '*bricolage*' (see Figure 3.a)—'creating a new artwork, concept, etc., by appropriating a diverse miscellany of existing materials or sources' (OED 2022)—and drawing on Barad's (2013: 18) notion of 'cutting together-apart…not separate consecutive activities, but a single event that is not one'—I created a new artefact from diffractive analysis of student artefacts, conversation transcripts and field notes. Student artefacts were produced at different times—however, inspired by Alexandra Lasczik Cutcher (2018) and Donna Carlyle (2020), the new artefact entangles these together through the notion of a '*pentimento*' (see Figure 3.b), whereby 'an earlier composition [is] seen through later layers of paint on a canvas' (OED 2022) and where authorship of the 'layers' is unclear. In this sense, the artefact is in a continual 'process of becoming' (Braidotti 2011: 41) and Barad's (2013: 28) notion of 'spacetimemattering' will be evoked, whereby 'neither

space nor time exist as determinate givens, as universals, outside of matter'; rather space, time and matter are entangled—'intra-actively configured and reconfigured in the ongoing materialization of phenomena'. However, drawing on the notion of agential cuts (section 2.3), a 'cut' of the artefact is presented for the purposes of this dissertation.

Figure 3.b: *The Arnolfini Portrait*.
Jan van Eyck, National Gallery, London 1434. Image from Wikimedia Commons, in public domain. Various alterations can be seen in infra-red reflectograms, including the faces and feet in the portrait.

3.2. Research design

3.2.1. Overview and planning diagrams

Figure 3.c and Figure 3.d detail the stages, activities and tasks involved in participant recruitment, data generation and analysis and presentation of findings, and these are referred to in the next sections. Larger versions (via a *Miro* board) can also be viewed here.

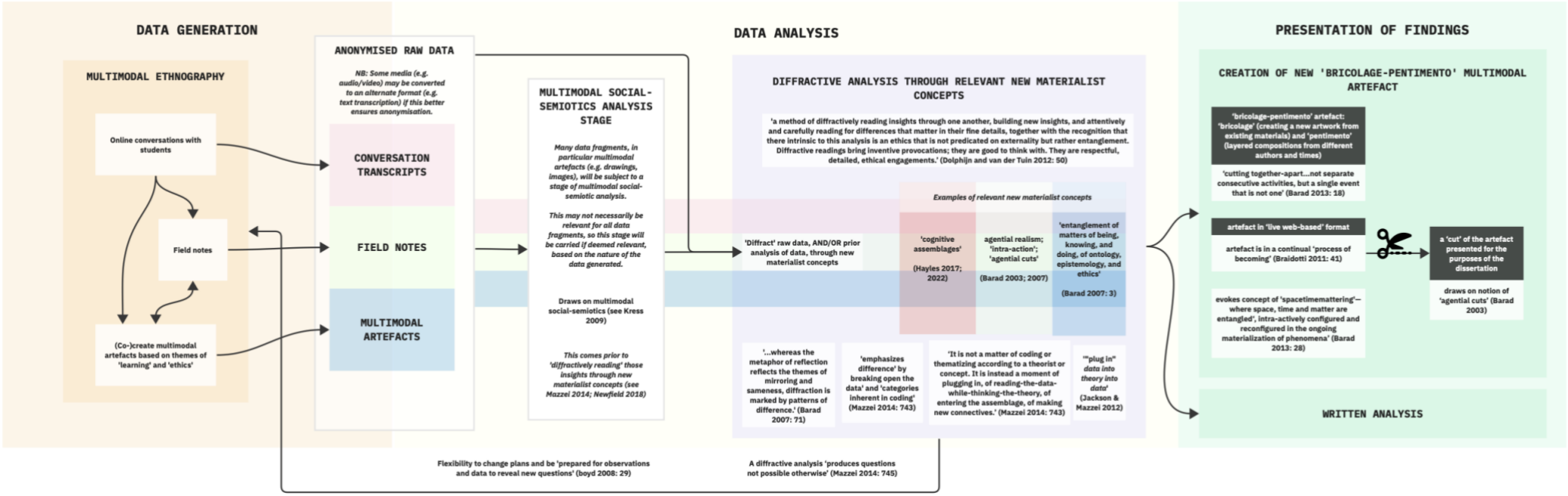

Figure 3.c: Overview of data generation and analysis, and presentation of findings.

View larger version in Miro board (also uploaded to submission dropbox).

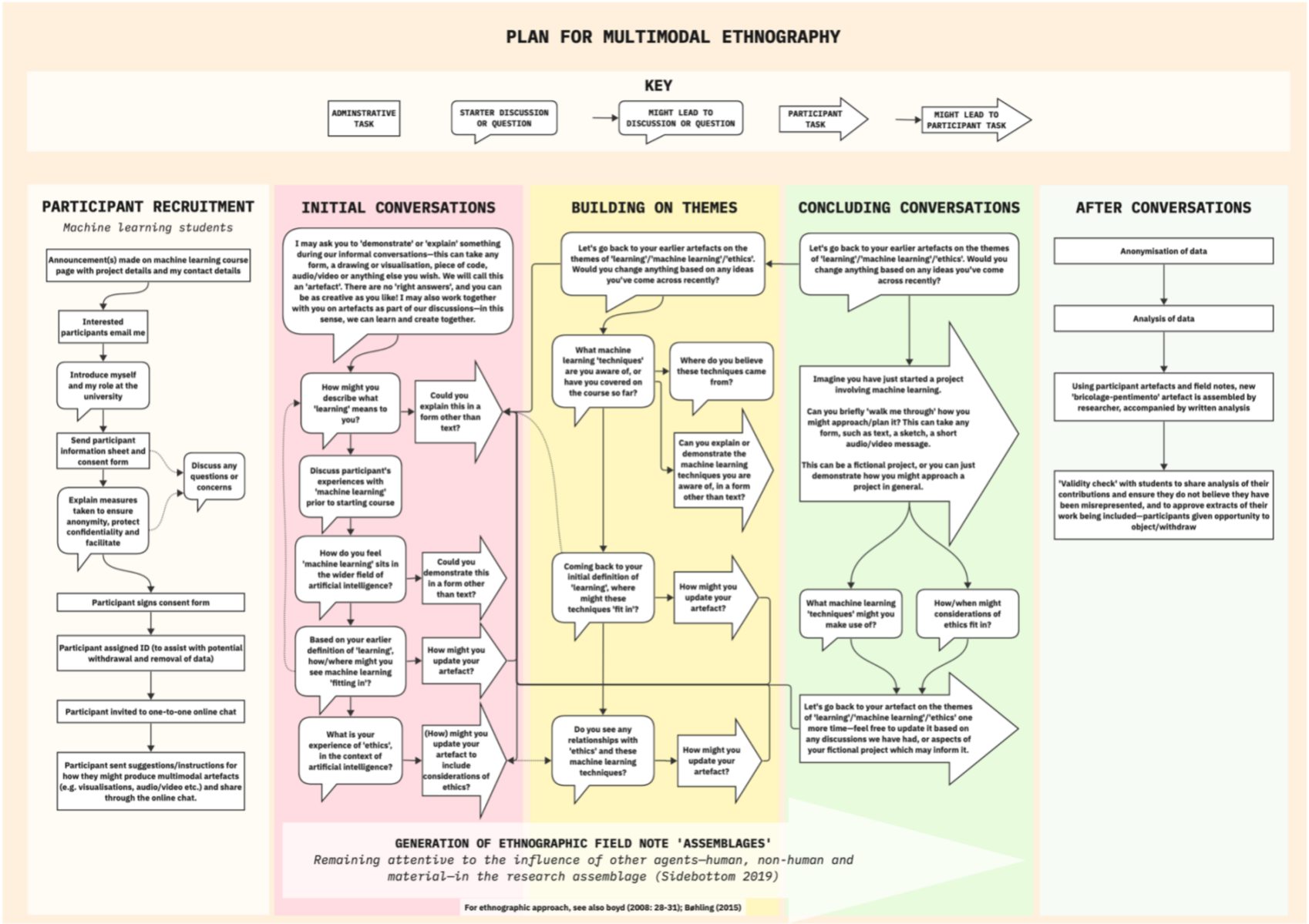

Figure 3.d: Plan for multimodal ethnography.

[View larger version in Miro board](#) (also uploaded to submission dropbox).

3.2.2. Participant recruitment

Participants were students recruited from several machine learning courses at TU Delft, a higher education institution where I worked as an educational advisor. Ethical approval was gained from the University of Edinburgh and TU Delft (paperwork has been uploaded to the ethics application dropbox).

Permission was sought from course instructors, and an announcement was distributed via the learning management system calling for volunteers. Students interested then contacted me directly, and were emailed an information/informed consent form (Appendix A1). The process following recruitment is detailed in Figure 3.d (under Participant recruitment).

Of the sixteen student participants who expressed interest, seven participants completed the study. As the project involved detailed multimodal asynchronous conversations which mostly took place over several months and involved relatively in-depth tasks, a large amount of rich data was produced even from this small sample size. Furthermore, the primary goal was not to produce generalisable data (although interesting insights were produced) but to demonstrate methodologically how multimodal and arts-based approaches may be combined with new materialist thinking through 'diffractive analysis' to explore how students learning about 'machine learning' conceptualise 'learning'. The data produced was subject to multiple forms of analysis (discussed below), some of which were methodologically experimental (reflexive discussion was a key aim).

Thus, given the project's aims and methodology—including in-depth discussion and multiple analysis stages—the sample size of seven participants was deemed appropriate. This is supported by literature on sample sizes, data saturation and qualitative research (see Creswell and Poth 2018; Hennink and Kaiser 2022). Moreover, a diffractive analysis aims to move beyond a conventional focus on searching for similarity within categories achieved through coding (Mazzei 2014). While engagement with differing theories leads to 'multiplicity, ambiguity, and incoherent subjectivity' (ibid.: 743), it reveals new questions not otherwise possible. This process is necessarily experimental, and articulation of it forms a significant part of discussion below and in chapter 4.

3.2.3. Data generation and analysis

Figure 3.c outlines data generation and analysis stages—multimodal ethnography, anonymisation of raw data, multimodal social-semiotic analysis, diffractive analysis and presentation of findings—and each stage is covered in the following sections (except the data anonymisation stage detailed in Appendix A2).

3.2.3.1. Multimodal ethnography

My ethnographic approach is informed by boyd (2008: 29) who, while acknowledging the importance of preparing questions for fieldwork, insists that we should not be bound by them; rather, 'researchers must be prepared for observations and data to reveal new questions'. Thus, while I planned the multimodal ethnography, I remained prepared for the data—and ongoing analysis—to open up new questions. Figure 3.d and Appendix A3 detail this semi-structured plan of conversations with participants, ongoing over several months, exploring themes of students' conceptualisations of 'learning' and 'ethics' through discussions and tasks including multimodal artefact generation. Secure online one-to-one discussions were necessary for students' confidentiality, and to comply with the ethical approval requirements.

It was provisionally considered that conversations should be carried out at the beginning, middle and end of participants' machine learning courses. However, it became apparent that students' availability for this was not possible. Moreover, seeking generalisable conclusions from a longitudinal study is not a goal; rather, it is to explore differences in conceptualisations of learning through differing theoretical lenses—multimodal social-semiotics and new materialisms—via an experimental diffractive methodology.

Inspired by Sarah Pink's (2011; 2015; 2021) ethnographic approaches, I envisioned conversations with students less as collecting data through interviews, and more as shared conversations 'through which new ways of knowing are produced' (Pink 2011: 271). I emphasised this to participants (Appendix A3), and employed collaborative multimodal techniques whereby artefacts were generated by participant and researcher, culminating in co-creation of the 'bricolage-pentimento' artefact (see below). Field notes were generated throughout the multimodal ethnography. Drawing on Frederik Bøhling's (2015) conceptualisation of 'field note' as 'assemblage', I blur subject/object and researcher/participant dualisms, and notions of authorship, by sharing field notes with participants and providing opportunity for input through a final "validity check" stage. A template for field notes/analysis is shown in Appendix A4, employed after anonymisation of raw data (Appendix A2).

3.2.4. Multimodal social-semiotic analysis

Once anonymised (see Appendix A2), data fragments were subject to multimodal social-semiotic analysis (see section 2.1). These included multimodal artefacts (such as drawings and images), but analysis was also informed by the wider social context revealed through text conversations. (See Appendix A4 for analysis template.)

3.2.5. Diffractive analysis

Following multimodal social-semiotic analysis, a *diffractive analysis* was performed. As laid out in chapter 2, the rationale behind employing a diffractive analysis is to bring insights from multiple approaches together—in this case, (human-centric) multimodal social-semiotics and anti-anthropocentric new materialisms. My approach is exemplified by Mazzei (2014), where data is threaded through different theoretical concepts to reveal new insights and questions. I diffracted the data through the new materialist concepts laid out in chapter 2—remaining attentive to affective flows between more-than-human entities within the research-assemblage (Fox and Alldred 2015; 2016), and considering 'intra-action' (Barad 2003; 2007) and the cognitive assemblage framework (Hayles 2017; 2022). This involved 'reading-the-data-while-thinking-the-theory' and 'making new connectives' (Mazzei 2014: 743), making notes alongside data fragments using my analysis template (Appendix A4).

3.2.6. Presentation of findings

I present my findings below in chapter 4 through written analysis, culminating in a new multimodal artefact. Informed by the aforementioned technique of *bricolage*, the multimodal artefact was assembled from a diffractive analysis of student artefacts, conversation transcripts and field notes generated during the multimodal ethnography stage. While data was generated at different times throughout the multimodal ethnography, the new 'bricolage-pentimento' multimodal artefact entangled these together through the notion of the aforementioned *pentimento*. As mentioned above, a 'cut' of the artefact is presented for the purposes of the dissertation—although theoretically it is continually becoming and never finished. Alternative possibilities for the presentation of this artefact arose through the experimental nature of the project; while ultimately out of scope, they are detailed in Appendix A6.

4. Presentation & Discussion of Findings

Following the research design detailed in chapter 3, this chapter presents a summary of the findings from data generation and analysis. As outlined there, my aim was not to produce generalisable conclusions—the small sample size does not make this feasible; and the diffractive methodology employed emphasises difference over sameness (Barad 2007). Rather, the goal of the study was twofold: first, to provide insights to the broad question *how do students studying machine learning conceptualise 'learning'?* from multiple perspectives—multimodal social-semiotics and new materialism—and reveal productive tensions (Kuby and Christ 2018); second, through an experimental methodology, to articulate how complex concepts such as the research-assemblage might be explored through modes beyond text.

Section 4.1 summarises the findings gained by exploring students' conceptualisations of learning from a multimodal social-semiotic perspective. Relevant quotations and multimodal artefacts from conversations are selected to demonstrate points (see Appendix A5 for all anonymised conversations, artefacts, field notes and analysis). Section 4.2 then summarises insights gained from diffracting this data through the new materialist concepts introduced in chapter 2. By following a rhizomatic and unpredictable journey, these explorations produced new questions and insights culminating in the production of the bricolage-pentimento artefact (Figure 4.h), a visual 'cut' of the research-assemblage. Through this, the complexity of more-than-human relations is made visible, conveying an honest account of the difficulty of locating where power lies in such an entanglement; discussion of this concludes this chapter.

4.1. Exploring conceptualisations of learning through a multimodal social-semiotic approach

According to Kress's (2009) multimodal social-semiotic approach set out in section 2.1, and in the context of the (multimodal) conversations I carried out with participants, attention should be paid to the meaning made through multiple modes and the wider environment in which this sign-making occurs. For example, why might participants represent their conceptualisations sometimes through text, and othertimes through drawing? What meaning is made through the choice of mode, the form of the drawing and the environment in which it was produced, and what cultural and semiotic resources were available to realise the drawing? Furthermore, signs are metaphors which frame thinking, often in ways that go unnoticed. This is discussed further—in the context of the metaphors revealed through analysis of the (multimodal) participant conversations—below.

While conversations often started with broad questions, such as 'how might you describe what "learning" means to you?', I am aware our conversations were framed within the context of a project exploring machine learning students' conceptualisations of (machine) learning and ethics. Therefore, their conceptualisations of learning may have been guided by the machine learning

context, although we discussed their perceptions of relationships between human and machine learning. Furthermore, the technical nature of setting up encrypted chats—although necessary for ethical approval, and to encourage participants to speak freely—may have framed the environment as somewhat technical. (These technologies and processes in fact feature later in the bricolage-pentimento artefact, as components entangled in the research-assemblage.)

While we communicated via text, participants were encouraged to represent conceptualisations through other modes. I sometimes produced visualisations of participants' text descriptions, and returned this to them with my analysis as part of the "validity check" and our collaborative 'shared conversation' (Pink 2011: 271). It is not possible to include below all multimodal artefacts produced—however, several key themes emerged through the multimodal social-semiotic analysis, and key data fragments have been selected for discussion below. Additional artefacts also inform the later diffractive analysis (section 4.2) and feature in the bricolage-pentimento artefact (Figure 4.h). (See Appendix A5 for field notes including all artefacts/analysis.)

4.1.1. Visibility of computational, mathematical and behaviourist tendencies

Several participants chose sketches to represent their conceptualisations of learning, which were quite mathematical and/or computational in nature—in the ideas they drew from and resources used (such as lined or graph paper, or flowchart/logic diagrams informed by mathematical concepts). One such example was the conceptualisation of learning drawn by one participant on lined paper, photographed and posted during our conversation (see Figure 4.a), and contextualised with this statement:

> **I tried to create this conceptualisation with as broad a view as possible. Asking myself questions such as: how do babies learn to walk? How do dogs learn to behave? And for me personally, How do I learn to drive a car? I think in all these types of situations this model could be applied.**

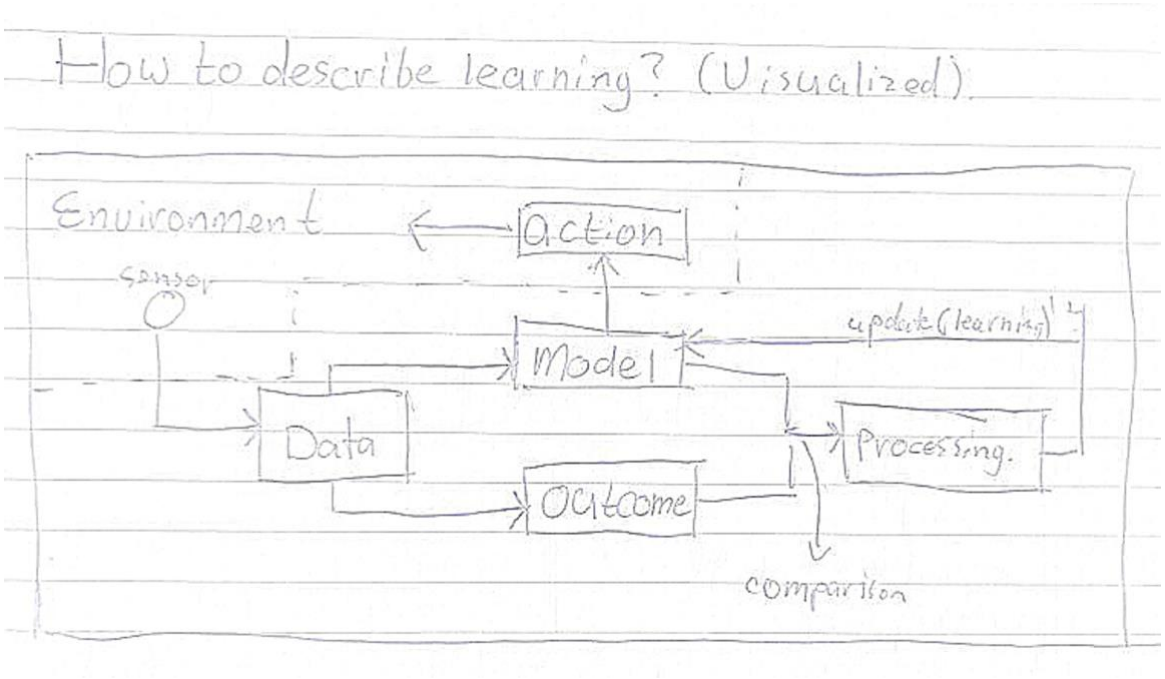

Figure 4.a: One students' drawing explaining how they would describe 'learning' in general.

While this conceptualisation was created to broadly describe 'learning' in general, rather than focusing specifically on human, animal or machine learning, the influence of the computational metaphor (discussed below)—and theories of learning such as behaviourism, rooted in animal psychology—is apparent. Interpreting this multimodal conceptualisation through Kress's (2009: 59) notion of 'signs-as-metaphors', the *signified*—'learning'—is represented by a process flowchart/diagram (the *signifier*) whereby a sensor captures data which, through a model, is processed. This processing updates the model (learns), resulting in actions to the (outside) environment. Such terminology—'actions' taken in an 'environment'—evokes language used in machine learning techniques such as reinforcement learning, whereby 'a learning agent must be able to sense the state of its environment to some extent and must be able to take actions that affect the state' (Sutton and Barto 2018: 31).

In her semiotic analysis of virtual learning environment human-computer interfaces, Siân Bayne (2008) explores what pedagogical and social practices are reflected through their visuality and spatial organisation, what meanings are produced, what versions of pedagogy are made visible and, crucially, what versions are made invisible. What theories of learning, then, are made (in)visible here? This conceptualisation of learning as a data-driven system—whereby learning is the result of such data being processed via a model—appears to emphasise computational metaphors which compare brain to computer and vice versa (Baria and Cross 2021).

The computational metaphor was visible in informing other participants' conceptualisations of learning, too. One participant opted for text to explain what learning meant to them:

> **To me learning is the process of making sense of data like text/speech or images/video and incorporating it into one's knowledge database…you don't just learn a sequence of letters but you extract the information from it and arrange it in your brain in a way that makes sense to you and that allows you to later on explain it to someone else.**

I later assembled an image during the analysis stage, which was also communicated back to the participant as part of the "validity check"—this is shown in Figure 4.b:

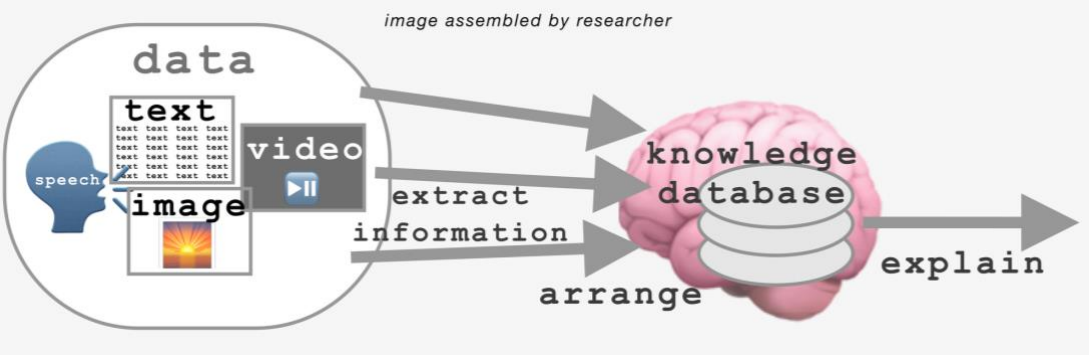

Figure 4.b: Image assembled by researcher conveying the 'knowledge database' described by a participant.

Here again the parallels with computational processes in this conceptualisation of learning is clear—meaning/information is 'extracted' from data (text, images, speech or video) and incorporated into a 'knowledge database'—signifying the brain. However, it should be noted that—according to a multimodal social-semiotic approach (Kress 2009)—it is important to pay attention to the wider environment. In this case, the computational environment through which conversations took place—where encrypted text, images and other media are stored in a database—is a part of the way meaning is made, as is the context of the research project focusing on machine learning and specifically involving machine learning students.

Why, though, is the aforementioned computational metaphor significant? According to George Lakoff and Mark Johnson (1980: 6), the 'human conceptual system is metaphorically structured and defined'—thus, the metaphor is not merely a matter of language or of words, (human) thought processes themselves are to a large degree metaphorical. Therefore, the computational metaphor has influence on how we understand brains and computers, and human relations with AI technologies (Baria and Cross 2021).

The appropriateness and impact of the computational metaphor is hotly debated in the fields of AI and neuroscience (Cobb 2020; Baria and Cross 2021). As Alexis Baria and Keith Cross (2021) argue, the inaccuracies of the computational metaphor might be visible to those with a deep knowledge of these fields but not necessarily to those without such a background. This lack of awareness, when such a metaphor becomes so familiar as to be accepted as literal, can be particularly dangerous (MacCormac 1984), limiting not only creativity in scientific inquiry but

reinforcing misleading ideas among the general public and even perpetuating (unintended) social and political messaging (Taylor and Dewsbury 2018). Metaphors have power, and are pervasive in shaping human cognition—oftentimes in way that are unnoticed and invisible to most (Lakoff and Johnson 1980; Lakoff 1992). As George Lakoff (1992: 481) remarks, 'what metaphor does is limit what we notice, highlight what we do see, and provide part of the inferential structure that we reason with'.

Furthermore, looking beyond the technical aspects of machine learning, what does the *notion* of (machine) learning do? What power lies in how (machine) learning is *conceptualised*? Exploring what he terms 'the social power of algorithms', David Beer (2017: 7) argues how the *notion* of the algorithm itself has power, suggesting that 'we look at the way that notions of the algorithm are evoked as a part of broader rationalities and ways of seeing the world'. Similarly, if we consider the *notion* of '(machine) learning', and 'artificial intelligence' more widely, what values and worldviews are promoted over others?

With this all in mind, and the prevalence of the (not necessarily intentional) computational metaphor in participants' conceptualisations of learning, what might be the implication of perceiving the world in a praxeomorphic way (see Bauman 1988)—that is, ascribing technological and computational characteristics to humans—in the context of (machine) learning and AI? Baria and Cross (2021) urge us to recognise the power of the computational metaphor, which can hide the complexities of the human brain while affording the computer more wisdom than is perhaps due, misleading and shaping our social behaviour in sometimes unintended ways through, for example, the marketing of 'intelligent' products, or the general practice and communication of science (see Taylor and Dewsbury 2018). They outline the importance of critical consideration of the social implications of AI development in the field of neuroscience, and more careful consideration of language as part of AI ethics courses in computer science programmes. Ultimately, they call for a new lexicon that opens up the possibility of new metaphors and conceptualisations being made visible and prominent.

Behaviourist theories, too, are made visible and perhaps privileged over alternative theories of learning. Similarities can be seen between the participants' conceptualisation of learning in general, and their drawing of reinforcement learning as part of their subsequent summary of machine learning techniques (shown in Figure 4.c, as drawn on graph paper). Reinforcement learning is a machine learning technique informed by behaviourist psychology (Knox et al. 2020), and rooted in studies of animal learning (Sutton and Barto 2018). Indeed, there is a reference to animal learning in one of the examples chosen by the participant to contextualise their general conceptualisation of learning (Figure 4.a)—'how do dogs learn to behave?'—which is perhaps tangentially reminiscent of classical conditioning, developed through physiologist Ivan Pavlov's (1902) experiments with dogs.

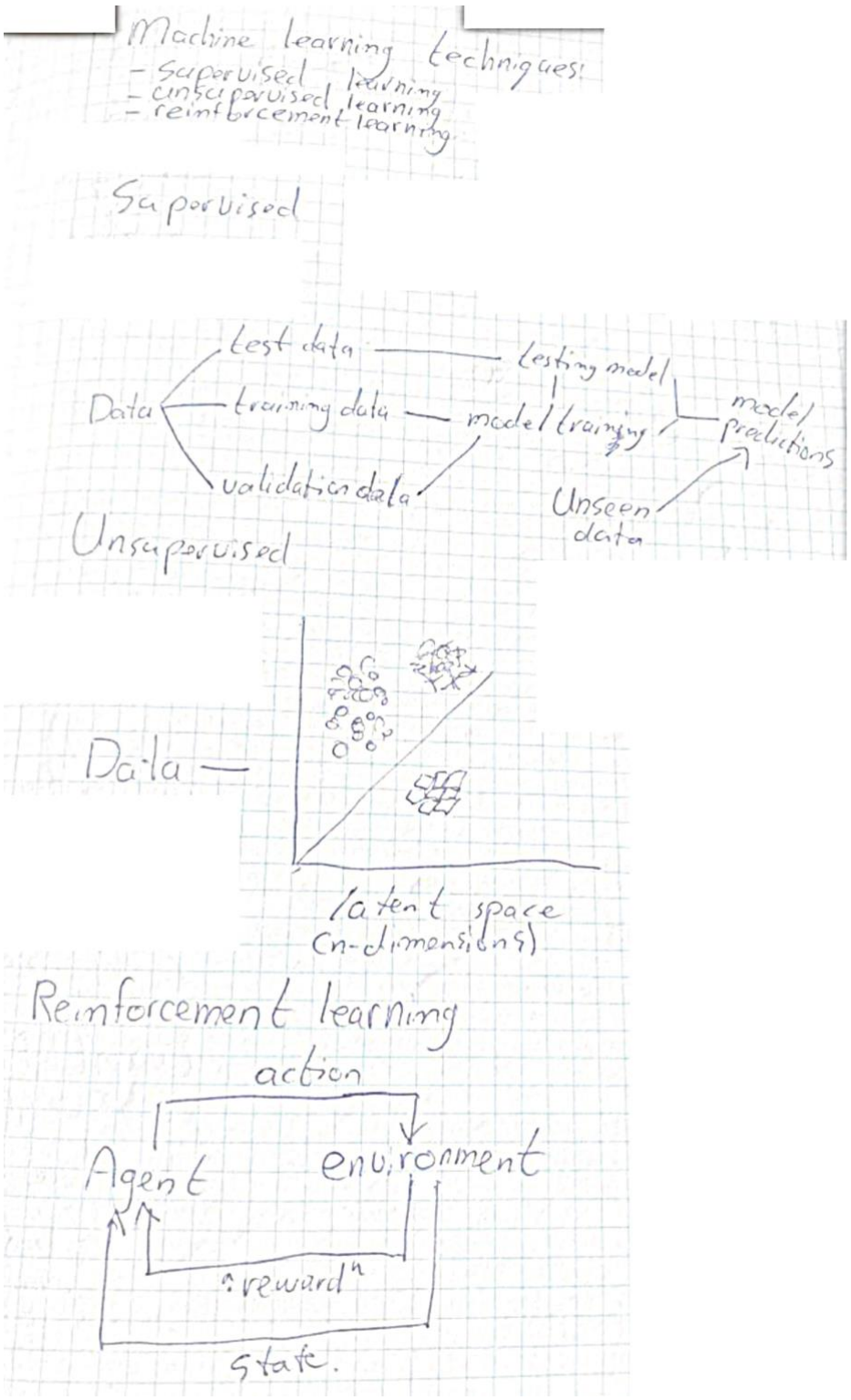

Figure 4.c: Participant's summary of machine learning techniques.

Figure 4.c's depiction of reinforcement learning broadly outlines what is known as the 'agent-environment interface' (Sutton and Barto 2018: 79), interacting in a Markov decision process

(named after mathematician Andrey Markov). Markov decision processes are 'a mathematically idealized form of the reinforcement learning problem', and intended to frame 'the problem of learning from interaction to achieve a goal' (ibid.). In this conceptualisation, the decision maker—or learner—is the *agent*, while everything outside this is the *environment*. The agent and environment continually interact, with the agent selecting *actions* and the environment responding with new situations and *rewards*—'numerical values that the agent seeks to maximise over time through its choice of actions' (ibid.). The diagram emphasises a model where the learning agent can be tightly controlled through a reward system, and the agent-environment dualism perhaps reinforces the notion that learning agents interact with, but are essentially separated from, the outside social environment; this sits in contrast with Barad's (2007) notion of intra-action.

The prevalence of Markov's ideas was further apparent in another participant's general conceptualisation of learning, as shown in Figure 4.d and contextualised with the following statement:

> **...learning is the update of states from one 'area' based on a change of states from another 'area'...I have a mental picture of molecules bouncing off each other as a way to visualize this updating. The mental picture of bouncing molecules gave me the incentive to discovered [*sic*] Bayesian updating.**

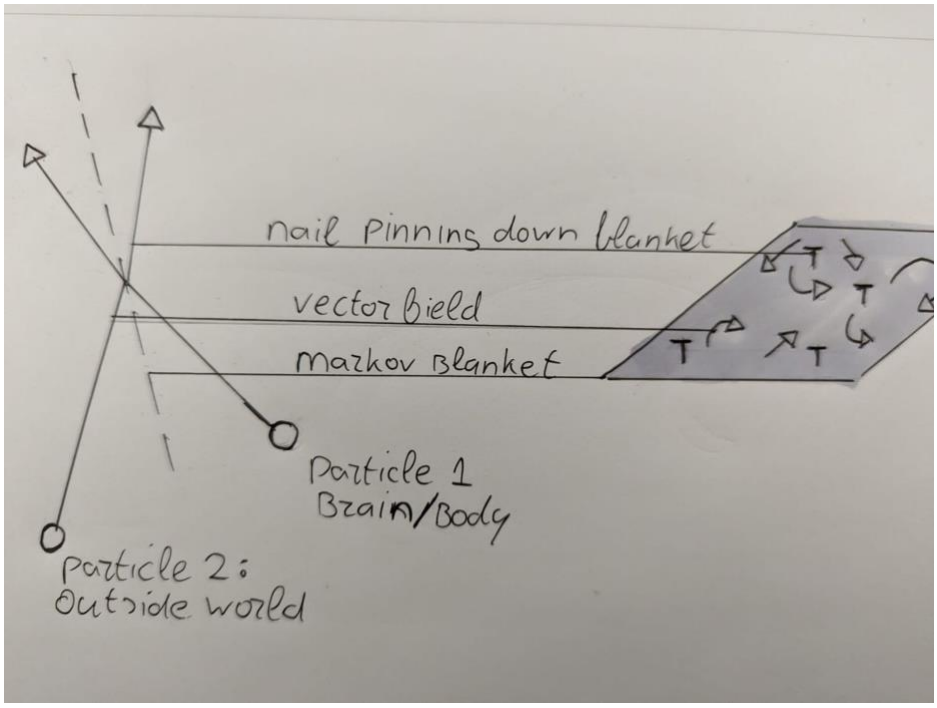

Figure 4.d: One students' drawing explaining what 'learning' means to them.

The influence of computational models of learning, such as Bayesian learning theory, to human cognition is apparent here, perhaps pointing to the way learning science has emerged from research in both AI and cognitive science disciplines (see Jacobs and Kruschke 2011; Wang et al.

2023). Furthermore, the 'brain/body' first 'particle' depicted is reminiscent of the Cartesian mind/body dualism, while the second 'particle' depicts the 'outside world' and reinforces the aforementioned agent-environment dualism.

While discussing further how this idea of 'bouncing molecules' might have emerged, the participant remarked it was 'almost straight from a high school physics book.' Perhaps their previous experiences with scientific educational materials plays a part in the emergence of this rather scientific and mathematical conceptualisation of learning, with links to Bayesian mathematical modelling and Markov decision processing theory (Jacobs and Kruschke 2011).

While some participants saw clear differences, there were parallels between several participants' general ideas of learning and AI/machine learning techniques. Having produced the below artefact (Figure 4.e), one participant later clarified their view of AI as mimicking the result of human thinking or behaviour which is 'desirable (rational)'—both evoking a 'biomimicry' of human thought processes (see Dicks 2016; Floreano and Mattiussi 2008), and the idea of a sometimes irrational human subject which AI technologies might improve upon.

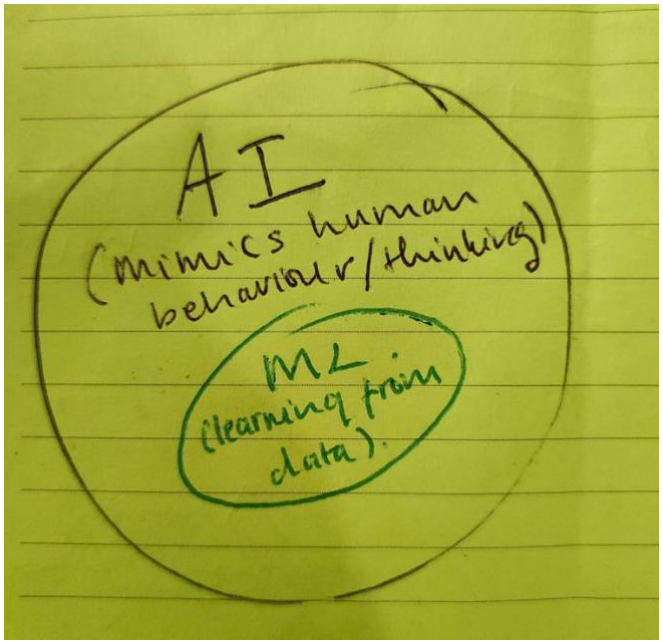

Figure 4.e: How one participant sees machine learning sitting within AI.

Another spoke of how they perceived similarities between human and 'artificial neural networks' (ANN):

> **…parallels can be made between the human neural network and ANN. Reinforcement learning and other techniques also use the same method we use to teach our children to not eat dirt for example. If a child gets in trouble for eating dirt, he might be less likely to do it again. The same thing goes for a model trying to find the solution to a problem that is being "punished" when he gives a stupid solution.**

The participant goes on to draw parallels between human and machine learning, framed by the notions of 'punishment' and 'reward':

> **Punishment and reward work the same way for learning. I think that I've even heard that reward-based learning works better (on human[s] that is). I think both method[s] are perfectly feasible for computers and can be brought together.**

The notion of 'punishment' and 'reward' make visible ideas of learning rooted in behaviourism (Skinner 1948; 1976 [1974]), appearing to link reinforcement learning with this conceptualisation of learning. This evokes B. F. Skinner's operant conditioning, and behavioural control through positive and negative reinforcement—although punishment is not equivalent to negative reinforcement (Papageorgi 2021). While the field of educational research has mostly moved away from behaviourism towards constructivist or cognitivist learning theories (Friesen 2018; Knox et al. 2020), behaviourist theories of learning—emphasising control of the passive learner—are brought to the forefront here. This arguably hides alternative ideas of the learner—such as constructivism's emphasis on active and autonomous individuals.

4.1.2. Conceptualising control and accountability/responsibility in (machine) learning systems

The notion of control was visible in other conceptualisations of learning. Another participant went on to explain one way they might conceptualise learning (for humans and machines); produced in the computer software *Paint*, it is shown in Figure 4.f and was contextualised with the following:

> **...showing someone exactly what needs to be done, like giving a detailed description of what is expected as a "framework"...a set of rules that bounds the space of exploration for learning.**

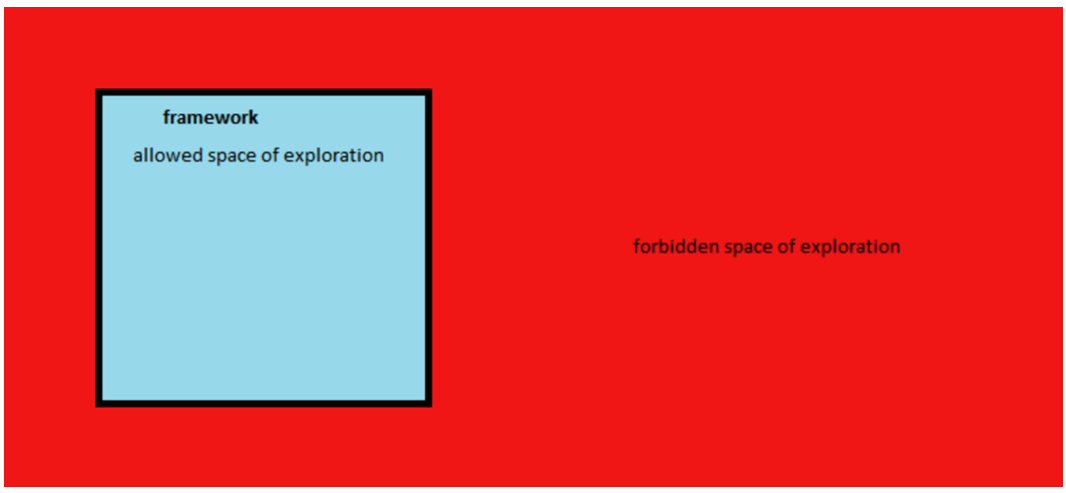

Figure 4.f: Participant's conceptualisation of learning, produced in 'Paint'.

This conceptualisation of learning appears to emphasise a system of control—a carefully bounded space determined by the rules of a framework, further elaborated by the participant:

> **…the system is giving a framework for the user…and the developpers [*sic*] are giving a framework for the system.**

Here, the learner (or user) is restricted to the 'allowed space of exploration', pre-determined by the framework of the system controlled by the developer. The *signified*—a framework for learning, with rules determining what can and cannot be explored—is represented by the *signifier* of a blue square (denoting what can be explored) within a red space (denoting what cannot be explored). Note the colour of the 'allowed space of exploration'—blue—and the 'forbidden space of exploration'. While vastly different meanings can be made through colour in different cultural contexts (Jacobs et al. 1991), in this context the red appears to be comparable to a stop sign—denoting a restricted area.

The notion of 'control' was a discussion point with other participants, too. One participant discussed the complexity of defining 'control' in the context of algorithmic systems, arguing that in general the machines' creators are the 'moral compass' and that the creators' values are reflected in the machine, but that this control is limited. Depending on the scenario, control might be distributed in different degrees between the machine, creator and users.

These conversations touching on how participants' conceptualised 'control' in the context of machine learning systems often led to discussing the ethical implications and notions of accountability and responsibility. One participant explained how they perceived relations between ethics and machine learning:

> **It seems to me that machines only perpetuate what humans input. If the input data is already bias/unrepresentative of the reality/unjust, the outcome will also [be] bias/unjust ("Garbage in, garbage out").**

This informed the production of the image shown in Figure 4.g, which I later assembled, and used to communicate my analysis with the participant as part of the "validity check":

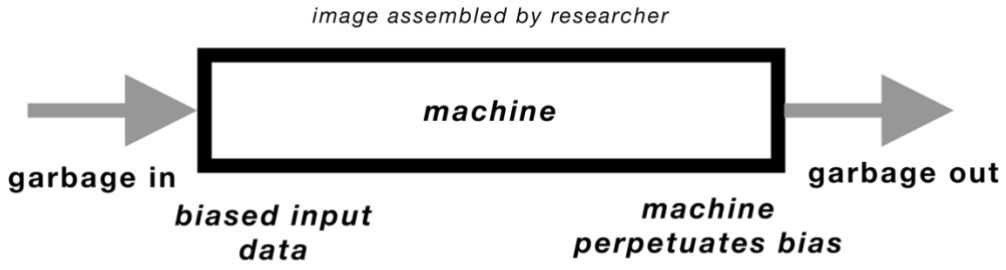

Figure 4.g: Image assembled by researcher conveying 'garbage in, garbage out' aphorism, inspired by a conversation about 'ethics' with a participant.

The aphorism ‘garbage in, garbage out’ is often employed in computing as a way to argue that ‘flawed input data or instructions will produce flawed outputs’ (Geiger et al. 2020: 325), and dates back decades—with a mention of the term appearing in a 1957 article in *The Hammond Times* (1957: 65), covering terminology used in military research carried out into ‘electronic’ or ‘mechanical’ brains by military mathematicians. This aphorism tends to emphasise a rather instrumentalist technological philosophy, whereby technology is seen as a ‘value-neutral tool’ (Borgmann 2009: 96) which has no control over the introduction of bias.

4.1.3. Summary of insights from multimodal social-semiotic analysis

A number of useful insights have emerged through this initial stage of multimodal social-semiotic analysis. There is an emphasis on computational and mathematical ideas of learning, and prominence of the computational metaphor (where brain is compared to computer). Specific theories of learning—in particular, behaviourism—and certain technological philosophies—such as instrumentalism—are also visible. Such computational, mathematical and/or behaviourist conceptualisations of learning tell only a partial picture (see Baria and Cross 2021; Dijkstra 1985; Langley 2011). However, reaching generalisable conclusions is not a key goal (see section 3.2.2); neither does this human-centred analysis tell the full picture. The following diffractive analysis seeks to look beyond an anthropocentric analysis by diffracting the above insights through different new materialist theories.

4.2. Diffracting the multimodal through new materialisms

Diffracting the above multimodal social-semiotic analysis through new materialist concepts allows us to go beyond focusing on the interests of the human subject, and ‘zoom out’ to consider the more-than-human entanglement of affective flows within the research-assemblage. However, given the complexity of such entanglements (and the apparatuses we build to study them)—which change with each intra-action (Barad 2007)—a diffractive approach can reveal more questions than answers. One such question is *how can we conceptualise agency in the context of this research-assemblage—through what affective flows have these conceptualisations of learning emerged?*

Exploring this question involved identifying ‘components’ and mapping these (and relations/flows connecting them) into different potential categories (based on different theoretical frameworks). This process was carried out both in table form and visually (culminating in the ‘bricolage-pentimento artefact’), each informing the other as they developed rhizomatically. This is discussed and presented below.

4.2.1. Components of the research-assemblage

This diffractive analysis seeks to move beyond representationalism, 'the view that the world is composed of individual entities with separately determinate properties' (Barad 2007: 55). If we are to move beyond exploring students' conceptualisations of (machine) learning through analysis of their multimodal *representations*, we instead consider a complex entanglement of intra-acting components—'traces of multiple practices of engagement' (ibid.: 53)—where agency resides in the affective flows between components. Each component is (re-)constituted through these ongoing intra-actions, and no components pre-existed nor came first. With this backdrop in mind, an initial 'cut' of components emerged, produced through boundary-making practices.

A selection of these components are shown in Table 4.a (full version in Appendix A7). They are categorised according to different frameworks—the research-assemblage (Fox and Alldred 2015) in column B; Hayles' (2017) cognitive assemblages framework in column C; and Hayles' (2022) hierarchy of attributes of ethical actors, drawing on Luciano Floridi and J. W. Sanders (2004), in column D. While the components in column A are separated for this initial categorisation, they in fact form a flow of affective relations (discussed below). This demonstrates my initial attempts to make meaning from the components emerging from a diffractive analysis, which informed the production of the 'bricolage-pentimento' artefact introduced next. Both are different agential cuts (see section 2.3), revealing different insights, questions and 'affective flows' (discussed later), and are included here in order to demonstrate the steps taken methodologically.

A: Component	B: Category	C: Hayles' (2017) cognitive assemblage framework	D: Hayles' (2022) hierarchy of attributes of ethical actors
conceptualisations of learning	events to be researched	mixed	
participants	participant	cognizers (actors)	(human) ethical actor (with degrees of capacity for interactivity, autonomy and adaptability)
researcher	researcher		
teachers	contextual elements		
participants' friends and family			
expectations/pressures on participants		mixed	
researcher's experiences with machine learning			
theories informing this research project		noncognizers (agents)	agents within interactive systems but with no moral accountability/responsibility alone
ethical approval processes/committees		mixed	
machine learning techniques			
modes/technologies/processes involved in participant conversations	research instruments/apparatus		
modes/technologies involved in presenting findings	recording and analysis technologies	noncognizers (although AI technologies complicate this)	agents within interactive systems but with no moral accountability/responsibility alone
technologies involved in drafting analysis			
course/teaching materials, methods and curricula	contextual elements		
other resources, e.g. YouTube videos, ChatGPT etc.			
participants' histories with machine learning			
(open source) coding communities (social/cultural norms)		mixed	
code involved in 'applying' machine learning techniques		noncognizer/agents—although programming technologies may act as cognitive supports	agents within interactive systems but with no moral accountability/responsibility alone
AI-related fiction		noncognizer/agents	

Table 4.a: Summarised list of intra-acting components in the research-assemblage.

Full version in Appendix A7.

4.2.2. Cutting conceptualisations together-apart—the assemblage of a 'bricolage-pentimento' artefact

From the components in Table 4.a, the bricolage-pentimento artefact in Figure 4.h (a larger version can be viewed here, via the web application *EasyZoom*) is the result of diffracting my prior analysis through new materialisms, producing something new. Conceptually, this is one 'cut' of the always evolving research-assemblage, produced to present insights from my explorations. As well as including artefacts and excerpts from participant conversations, additional imagery conveys points made by participants through text, and blue two-way arrows convey affective flows between components.

As introduced in section 3.1.2, the *bricolage* technique combines components of the research-assemblage together through boundary-making practices—these are one 'cut' of something always changing—evoking an entanglement of space, time and matter and Barad's (2013) notion of 'spacetimemattering'. Image transparency is employed, allowing components to be seen through one another, and evokes notions of *diffraction* (waves/ideas interfere and produce something new) and the *pentimento* (compositions from different authors/times are seen through one another). This brings into question authorship/agency of components, and brings to the forefront the performative affective flows drawn as arrows between components of the research-assemblage. The Creative Commons licence under which the artefact is released speaks to the remixing of components and breaking apart of individual authorship.

The rationale behind this artefact is twofold: first, to make visible insights gained from complex new materialist concepts in an alternative multimodal way, rather than relying on dense text; second, the bricolage—remixing and reworking existing multimodal conceptualisations with my own—speaks to my own entanglement within the research-assemblage. This artefact experimentally articulates one way to bring multimodal and new materialist perspectives together productively. Furthermore, the diffractive approach employed opens up the possibility of 'a true transdisciplinarity' through detailed and attentive readings of different ideas through one another, generating new and inventive insights not possible when approaches and disciplines are pitted against each other (Bozalek and Zembylas 2018: 51). It is hoped this may inspire development of creative multimodal educational activities in transdisciplinary AI programmes which seek to engage with a complex multiplicity of perspectives. However, it should be noted that accessibility and inclusivity should be carefully considered during the educational design (see Appendix A6).

Although, the artefact is presented in a 'static' two-dimensional format (due to limited scope, and for discussion/assessment purposes), future research might explore the constantly changing research-assemblage by drawing on 'live sociology' (Back 2012) and 'live methods' for 'real-time' investigation/presentation (Back and Puwar 2012). Alternative possibilities include 'modelling' the research-assemblage in a three-dimensional, animated and/or 'interactive' format, albeit with

methodological, technical, accessibility and ethical challenges. Several possibilities are mentioned in Appendix A6, which may inspire development of educational activities where students and teachers creatively and critically rethink conceptualisations of (machine) learning and entanglements with ethics.

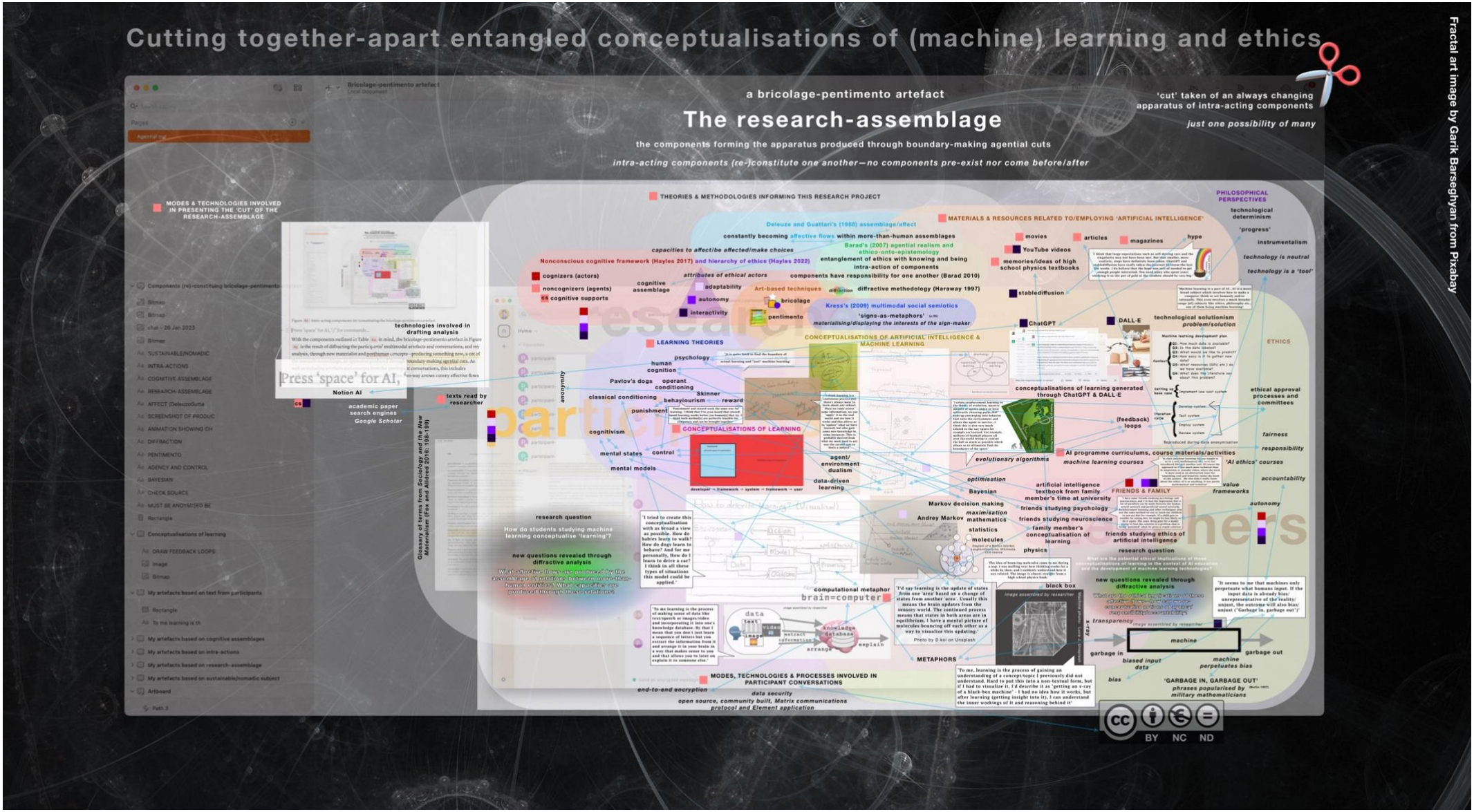

Figure 4.h: Intra-acting components (re-)constituting the bricolage-pentimento artefact.

View larger version via EasyZoom application (PDF and image versions also uploaded to submission dropbox).

4.2.3. Zooming out to take in the wider entanglement of affective flows

Figure 4.h visualises a cut of the research-assemblage and the affective flows—producing the capacity for intra-action (Barad 2007)—within. Through this dense complex visualisation, the focus for analysis is (visually and methodologically) "zoomed out" from individual components to take in the affective flows between components and the capacities which they produce (Fox and Alldred 2015). Furthermore, by diffracting my initial research questions through Fox and Alldred's (2016) new materialist sociologist framework (see chapter 2), new questions about these affective flows were revealed (shown in Table 4.b).

A: Initial research questions	B: New questions revealed through diffractive analysis
How do students studying machine learning conceptualise 'learning'?	What affective flows are produced by the assemblage of relations between more-than-human entities? What capacities are produced through these relations?
What are the potential ethical implications of these conceptualisations of learning in the context of AI education and the development of machine learning technologies?	What are the ethical implications of these affective flows—how can we re-conceptualise notions of agency/responsibility/accountability?

Table 4.b: New questions revealed through diffractive analysis.

There are many (always changing) affective flows—many sides of the assemblage—which could be explored. However, through diffractive analysis and production of the bricolage-pentimento artefact, several affective flows emerged in particular—these are summarised in Table 4.c (full version in Appendix A8). Each have capacities to produce events associated with the becoming of the conceptualisations of learning and associated ethical effects. These flows connect a diverse range of more-than-human entities—none privileged over another. These are listed in column B—in no particular order—in a 'cloud' of intra-acting relations (Fox and Alldred 2016: 30). While the affective flows in Table 4.c are separated/categorised for clarity, they are in fact entangled within the research-assemblage (through which conceptualisations of learning emerge).

A: Flow	B: Connecting
'Computational'	student — experience with computers/data processing — machine learning applications experienced/developed by students — flowchart/logic diagrams — computational/black box metaphors
'Mathematical'	student — squared graph paper — Markov blanket — Bayesian models — maximisation — optimisation — statistics
'Behaviourist'	student — behaviourist learning theories — punishment/reward — control — reinforcement learning — agent/environment dualism
'Personal experiences'	student — friends studying ethics/neuroscience/psychology — family's textbooks — personal/family members' ideas of learning
'(Institutional) education'	student — teachers — educational programmes — materials/methods/activities — educational philosophies — institutional structures/ initiatives — administrators
'Philosophical perspectives'	student — beliefs about 'truth' and (non-)neutrality of technology — technological determinism — notion of progress — technological solutionism — instrumentalism
'Ethics'	student — beliefs about 'ethics' and 'bias' — 'garbage in, garbage out' aphorism — AI ethics courses — friends studying ethics — notions of control/responsibility/accountability/autonomy
'Media'	student — YouTube videos — movies — articles — hype
'Research project'	student — researcher — theories/methodologies — technologies/processes — ethical approval processes

Table 4.c: Summary of affective flows identified through diffractive analysis.
Full version in Appendix A8.

Some themes (such as computational/mathematical tendencies) have already surfaced through the prior multimodal social-semiotic analysis. However, diffracting this through new materialisms shifts the focus of power from human participants to affects deriving from more-than-human relations which create the conditions through which these conceptualisations of learning can emerge. By 'ascribing "affective" capacity' to all kinds of matter' (Fox and Alldred 2016: 30), we focus instead on the complex connection of relations and human/nonhuman boundaries are destabilised (Barad 2003).

These relations include connections to past and present personal experiences—for example, friends studying neuroscience or ethics, or a family member's idea of how a subject might be 'correctly' learned ('personal experiences' flow). These relations are themselves entangled with resources and experiences related to educational institutions—such as a physics textbook from high school, (or an impression of what one might be like), or machine learning courses materials and classes they have experienced ('(institutional) education' flow).

These intersect with 'computational' and 'mathematical' flows—through computational metaphors, flowchart/logic diagrams and graph paper employed in participants' artefacts, statistical/probabilistic theories and Bayesian models, and many other relations. Such flows made visible specific computational/mathematical conceptualisations of learning, as 'behaviourist' flows made behaviourist theories of learning visible—through notions of punishment, reward and control, experiences with machine learning techniques such as reinforcement learning and notions of strict 'frameworks of exploration' for users. These flows all intersect with notions of control, responsibility and bias, technological philosophies and beliefs about the (non-)neutrality of technology in the 'ethics' and 'philosophical perspectives' flows, all themselves entangled with the 'media' flow—with YouTube videos, movies and hype. Finally, these flows are all connected to me as researcher, and the theories, methodologies, technologies, ethical approval processes as part of the 'research project' flow.

What, then, are the *capacities* produced through such relations? These are complex questions, not least given the constantly changing nature of the affective flows within the research-assemblage, a portrayal of one cut—one possibility—of which has been attempted here. As Hayles (2022: 2) asserts, 'the point of asking "Who's in charge here?" is to indicate the impossibility of any simple answer and the urgent necessity to address the ethicopolitical complexities and ambiguities the question implies'—embracing this complexity is difficult but necessary.

A diffractive approach has demonstrated a modest attempt to address these complex issues from different perspectives, revealing tensions but productive questions. Table 4.a categorised components according to Hayles' (2022) cognitive assemblage framework, and this somewhat differentiated between human ethical actors (attributing capacities for interactivity, autonomy and adaptability) and nonhuman agents (with sometimes lesser capacities)—in tension with other new materialisms making no such distinction (see Fox and Alldred 2016). In the context of Barad's (2007) intra-acting components—where none are privileged over another—*how much attention should be paid to individual components?* However, if we focus only on intra-actions *between* humans, nonhumans and material components, *are efforts to make change associated with the lives of humans futile?*

Fox and Alldred (2016) argue efforts for change should not be abandoned, but re-thinking our conceptualisations of power relations shaping affective flows is necessary to address important issues with the world. Taking a non-anthropocentric perspective, and rejecting delineation of humans/nonhumans, is not 'some arbitrary hang-up over dualisms' (ibid.: 44); rather, anthropocentric dualisms have been founded upon politics of colonialism, racism and sexism (Haraway 1988; Braidotti 2013). Anthropocentric approaches tend to locate conceptions of morality within human individuals, while ethics is positioned within abstract discourses of universal human rights—obfuscating human responsibility (Taylor 2018). Thus, far from complicating or

restricting change, moving beyond a humanist stance—and its individual/universal dichotomy—in fact serves to make visible ethical issues for which humans have failed to take responsibility.

Taking a non-anthropocentric approach, then, to interrogate affective flows which create the conditions through which conceptualisations of learning can emerge is necessary to enable change. This includes reworking how (machine) 'learning' is framed in, and ethics integrated into, transdisciplinary AI programmes—in order to contextualise socio-political histories of learning theories involved in machine learning techniques, and to make visible a wider range of ideas about learning. Such change also includes critical consideration of the complex and diverse range of affective flows through which learning is conceptualised. This project's diffractive approach demonstrates one productive way this might be realised, by engaging with multiple approaches to gain fresh insights and directions for future exploration.

While the points made so far have focused on the wider entanglement of affective flows within the research-assemblage (shown in Figure 4.h), the following section focuses "zooms in" on a selection of data fragments to demonstrate the questions and insights a diffractive analysis can yield.

4.2.4. Zooming back in to diffract multimodal data fragments through new materialist concepts

In this section, I focus on several data fragments analysed through a multimodal social-semiotic lens above, and take a fresh look by diffractively reading them through new materialist concepts. Due to the project's scope, it is not possible to cover all data fragments here, however the below have been selected to demonstrate key insights gained.

To achieve this aim, the images have been annotated with a 'layer' of overlapping rhizomatic affective flows, produced through diffractive analysis. These images have also been added to the aforementioned bricolage-pentimento artefact in the *EasyZoom* application, viewable as clickable annotations. These annotations are just one possibility—one 'cut' produced through boundary-making practices—in an entanglement of always evolving affective flows. The first of these is shown in Figure 4.i (adapted from Figure 4.a).

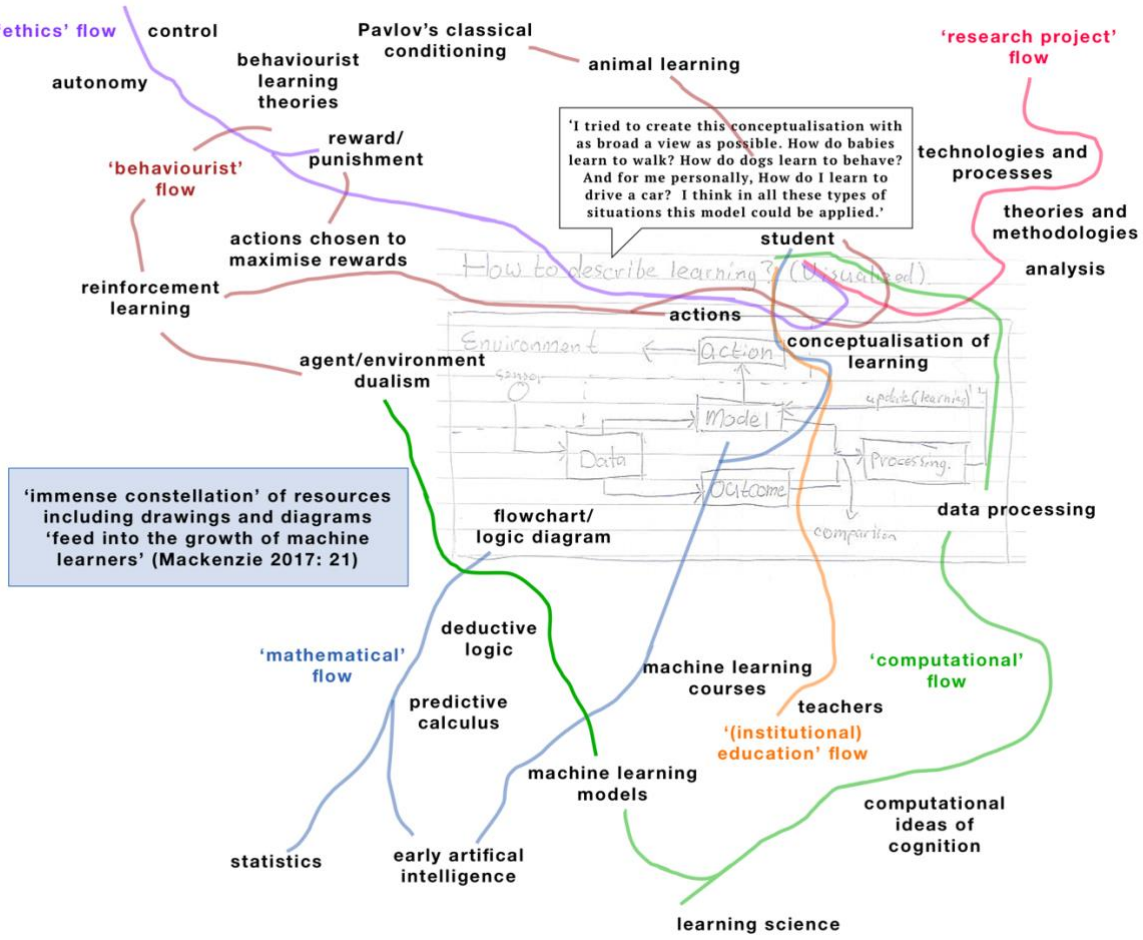

Figure 4.i: Diffraction of participants' conceptualisation of learning through new materialist concepts.
Reveals intra-acting affective flows, creating the conditions possible for conceptualisations to emerge.
Adapted from Figure 4.a.

The prior multimodal social-semiotic analysis pointed to the influence of the computational metaphor on this conceptualisation. However, drawing on Mazzei (2014), to read this specific data fragment diffractively through new materialist concepts—specifically the aforementioned notions of affect (Deleuze and Guattari 1988), affective flows within assemblages (Fox and Alldred 2016) and Barad's (2007) intra-action—prompts the following questions:

- *How does this participant intra-act with human, nonhuman and material entities in ways that produce the becomings of these conceptualisations of learning?*

- *From where do affects derive from—for example, ideas and experiences—which create the conditions where it is possible for these conceptualisations to emerge?*

Such questions de-privilege individual agency from the participant—contrasting with the multimodal social-semiotic analysis—focusing instead on the affective flows the participant is

caught up in and which make this particular conceptualisation of learning possible. Figure 4.i. makes visible 'computational', 'mathematical' and 'behaviourist' flows—involving computational ideas of cognition, ideas about animal learning and specific statistical logic flow diagram stuctures (see Mackenzie 2017)—to name just several intra-acting more-than human components. These affective flows produce the conditions whereby a very specific conceptualisation of learning emerges, mathematical and computational in nature, and controlled through data-intensive 'logical' and 'rational' systems.

Some of these insights have commonalities with those gained through multimodal social-semiotic analysis—conceptually the increased intensity caused by 'constructive interference' of two waves (see section 2.5). However, Barad (2007: 205) emphasises our mutual accountability/responsibility to 'perpetually contest and rework the boundaries' intra-actively co-constituted between humans/nonhumans which enact certain exclusions. For this project, the narrow focus of multimodal social-semiotics, locating agency primarily in human participants, could be seen as one such exclusion. A deconstructive analysis—performed by diffracting through new materialist ideas—reworks conceptualisations of agency into entanglements of affective flows; making these visible transforms our consideration of ethics (discussed below).

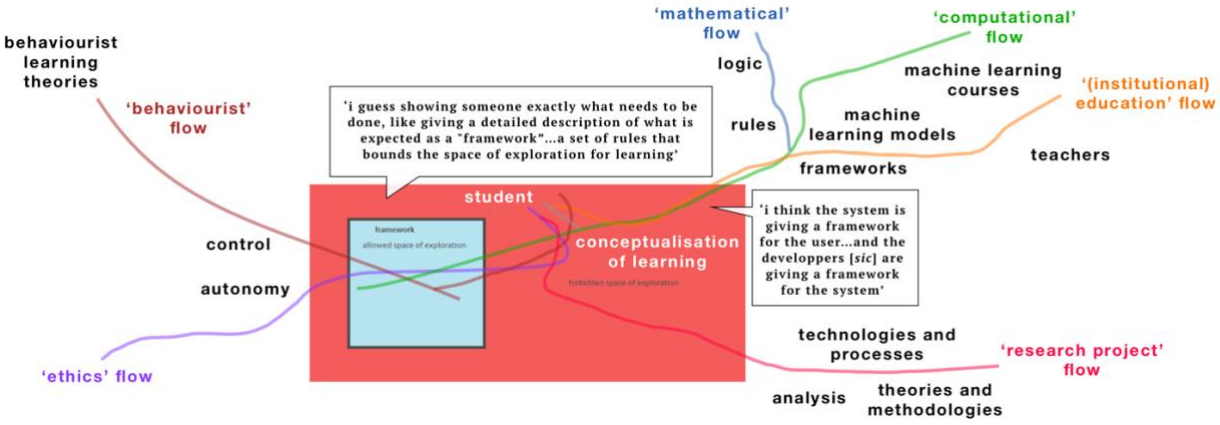

Figure 4.j: Diffraction of another participants' conceptualisation of learning.
Adapted from Figure 4.f.

A second data fragment has been annotated with another 'cut' of affective flows, shown in Figure 4.j. While a multimodal social-semiotic analysis pointed to a conceptualisation of learning based on a carefully controlled rule-based framework, reading this data fragment diffractively through new materialist concepts—such as Barad's (2007) ethico-onto-epistem-ology and Hayles' (2017; 2022) cognitive assemblages framework and hierarchy of ethical actors—prompts the questions:

- *What entanglement of intra-acting agents and ethical actors—with potentially differing capacities for making moral choices—are producing the conditions possible for a conceptualisation of learning based on control, rather than student autonomy?*

- *If making visible ideas of learning emphasising student autonomy is an ethical goal, how/who/what might be responsible/accountable for such a change?*

- *If deep integration of ethics into machine learning courses—including critical consideration of histories of (machine) learning theories—is an ethical goal, how/who is responsible/accountable for such an endeavour?*

Shifting our focus from human students and teachers involved in AI education to these affective flows transforms consideration of the origins for such conceptualisations of learning—with ethical implications for AI technologies students may later be involved with. A different combination of affective flows would make possible different conceptualisations of learning, yet introducing 'alternative' conceptualisations within AI education could not be enacted by human students and teachers alone. Rather, change would be enacted through a reworking of boundaries co-constituted through intra-acting human/nonhuman components—for which (human) students and teachers would be mutually accountable, together with other nonhuman components (highlighted in Table 4.a). Such change would be continuous and complex—pedagogic events are multiplicities of iterations and keep making connections (Taylor 2018).

It is arguable how one might conceptualise sharing/allocating the accountability/responsibility for change between human/nonhuman components. While a diffractive methodology may expose tensions between approaches as to whether allocating differing levels of responsibility to components is possible, it raises productive questions for critical discussion—which could feature within transdisciplinary AI educational activities. For example, drawing on Floridi and Sanders (2004), Hayles (2022: 15-16) incorporates moral accountability and responsibility into her cognitive assemblage framework, whereby 'morally responsible agents' are 'responsible for creating the conditions within which accountable agents function'—albeit with a 'spectrum of possibilities in between' accountability and responsibility determined by the 'potential for ethical action' agents may have. One educational activity might involve 'mapping' accountability/responsibility to the agents/components in the bricolage-pentimento artefact (Figure 4.h), serving as a starting point for critical discussion (see Appendix A6).

4.2.5. Summary of insights from diffractive analysis

Reading these data fragments diffractively through new materialisms, one cannot simply say that learning theories based on computational ideas of cognition, or behaviourist theories, solely influenced these students' conceptualisations of learning. Nor can one simply assert that these conceptualisations of learning are representations of ideas from their machine learning studies, or

that their studies alone changes the nature of machine learning applications they may in future develop. Rather, these (and many other) aspects co-constitute one another through a complex array of constantly changing affective flows.

A key insight gained is that embracing this complexity is challenging but necessary to reveal an honest account of the entanglement of affective relations making possible conceptualisations of learning, with wide-ranging ethical implications. Differing approaches offer useful insights, sometimes at tension, but productive in the questions they generate for critical discussion—discussion important to integrate into transdisciplinary AI educational activities (see Appendix A6). Taking into account the affective flows highlighted through diffractive analysis transforms consideration of these conceptualisations of learning, their origins and wider impact. This changes, too, consideration of how 'alternative' ideas of learning—and historical contextualisation of dominant learning theories in machine learning—might be introduced into transdisciplinary AI education.

Considering Barad's (2007) ethico-onto-epistem-ological stance, the changing possibilities for these intra-acting affective flows producing the becoming of these conceptualisations of learning are intertwined with ethics. Responsibility does not lie solely with students, teachers or curriculums, yet paying attention to the phenomena produced through these intra-acting components is an important if complex ethical matter. As Barad (2007) has recast interactions as *intra-actions*—questioning independence/boundaries of individual subjects—Carol A. Taylor (2018: 90) calls for posthuman/new materialist ethical interventions to be recast as *intra-ventions* involving an 'enactment-in-relations amongst all bodies, and not as a "thing" possessed by a sovereign and boundaried human subject which can be deployed "on" or "towards" "others" as if "they" were somehow "outside" the self'. This draws upon Haraway's (2016) 'response-ability', which re-orientates ethical practices around *reciprocal obligations* emerging through human/nonhuman meetings.

Thus, change cannot come from teachers nor students alone, nor from one discipline, initiative or curriculum—accountability/responsibility is mutually co-constituted and reciprocal between all humans, nonhumans, material and abstract entities. A collaborative reworking of boundaries made through the wider context of more-than-human components is important—ideas introduced to them at school, by friends and family, by the media and so on. Institutional educational initiatives to integrate ethics into AI programmes/machine learning courses can only go so far; however, that is not to dismiss the possibility for change, which could contribute to creating the conditions possible for conceptualisations of learning to emphasise, for example, student autonomy rather than notions of control.

While this diffractive analysis has perhaps raised more questions than answers, given its limited scope it has been beneficial in making visible limitations of centring human 'free will' in anthropocentric ethical frameworks, and instead highlighting the importance of interrogating the

diverse range of affective flows within assemblages. This applies to conceptualisations of learning, but also to connections with AI education and machine learning applications later developed by students; where AI/machine learning becomes involved with educational practice, such affective flows may even loop back into education. The intra-actions through which these components emerge are deeply ethical, and it is hoped this analysis has opened up new directions for exploring these complex issues.

5. Conclusions

The production of the artefact and the surrounding analysis aligns with the two broad goals laid out in chapters 1 and 3. First, useful insights were gained into students' conceptualisations of learning from different perspectives—primarily multimodal social-semiotics and new materialisms. Second, it has been methodologically demonstrated how one might employ a diffractive methodology to bring together these different perspectives and produce *new* questions and insights with a renewed focus. Furthermore, the multimodal arts-based practices employed in producing the artefacts may inspire new ways to critically consider complex more-than-human entanglements, in research as well as AI education.

From the initial multimodal social-semiotic analysis, students' conceptualisations of learning broadly tended to emphasise computational and mathematical ideas of learning, privilege behaviourist theories of learning and reinforce instrumentalist philosophies of technology. Within the project's limited scope, such an emphasis may partly appear to point towards ideas of learning shaped by a combination of behaviourism and data-driven machine learning technologies. However, the primary goal of this project was not to produce generalisable conclusions, which would be problematic given the small sample size, and single institution from which participants were recruited. Rather, I sought to move beyond conventional approaches and engage experimentally with a diffractive methodology to affirmatively bring together apparently incompatible theories—such as multimodal social-semiotics and new materialisms. A diffractive methodology searches for 'patterns of difference' (Barad 2007: 71) rather than similarity within categories or generalisable conclusions (Mazzei 2014) and, as demonstrated through this project, it offers productive ways to engage with (human-centred) multimodal social-semiotics and anti-anthropocentric new materialist concepts such as intra-action, (cognitive) assemblages and affect(ive flows).

A diffractive approach sets up a relationship whereby ideas are not opposed, but are co-constituted and produce new insights and questions (Mazzei 2014)—and, in the case of this project, a bricolage-pentimento artefact (Figure 4.h) conveying the complexity of the research-assemblage and co-constitution of ideas. Though this departure from conventional methodologies produces ambiguity and multiplicity (Mazzei 2014), it ultimately aims to enrich our understanding through 'mutually informative insights' (Barad 2007: 208). The constructive/deconstructive analysis carried out has affirmatively engaged with different theories to take a fresh look at conceptualisations of (machine) learning and their potential ethical implications. Crucially, the focus of enquiry has been shifted from a human-centric analysis of multimodal representations of conceptualisations of learning produced by students of machine learning to the affective flows between more-than-human components through which these conceptualisations emerge. While this complicates the question *'who is accountable/responsible for these conceptualisations of*

learning and their ethical implications?', if one honestly engages with the entangled complexity of more-than-human relations, a simple answer is impossible; yet, this only demonstrates the urgent need to address such ethical questions (Hayles 2022) by developing new approaches. Through the experimental methodology carried out in this project, modest attempts have been made to address such questions through multimodal and diffractive approaches. These approaches might offer creative ways for AI educational programmes to engage with complex ethical questions and notions of accountability/responsibility in the context of more-than-human entanglements (see Appendix A6).

While conventional AI ethics courses are limited due to their lack of deep integration of questions of ethics into core machine learning courses (Lim et al. 2023; Saltz et al. 2019), diffractive approaches open up the possibility of 'a true transdisciplinarity' (Bozalek and Zembylas 2018: 51). Such transdisciplinarity is realised by avoiding the pitting of one approach or framework against another and instead, as Barad puts it, 'diffractively reading insights through one another, building new insights, and attentively and carefully reading for differences that matter in their fine details, together with the recognition that there intrinsic to this analysis is an ethics that is not predicated on externality but rather entanglement' (Dolphijn and van der Tuin 2012: 50).

It is hoped that the bringing together of arts-based multimodal social-semiotic approaches with new materialisms might offer practical inspiration for truly integrating ethics throughout machine learning through creative transdisciplinary educational activities. Future research might build upon the creative methods employed in this project—and new materialist theories—to design activities which involve collaborative and critical reflection on how conceptualisations of (machine) learning might be framed ethically. Additionally, through the open-minded rhizomatic nature of this project, future possibilities for such collaborative activities were identified which could explore 'live methods' for 'real-time' investigation and presentation (Back and Puwar 2012). While exploring them further was outside the limited scope of this project, they are detailed in Appendix A6, along with important accessibility and inclusivity considerations for implementation of any educational activities of this kind. It is hoped such collaborative activities may offer a way for those from differing disciplines and theoretical backgrounds to come together and engage productively with a diverse range of frameworks for considering ethical issues surrounding AI and machine learning.

Through a diffractive methodology this project opens up the possibility of productive discussion around how one might approach the development of transdisciplinary AI education programmes where, for example, ethics is truly integrated with machine learning. How should decisions related to the design of such education be made, and how should such educational decisions be critically analysed? If one diffractively reads such decisions through new materialist ideas, considering *distributed* decision-making within complex assemblages of constantly becoming affective flows, this shifts the focus from human teachers and students alone. This may open up nuanced critical (potentially multimodal) discussion and analysis interrogating which conceptualisations of learning

are being brought to the forefront and which are hidden in AI education—and what ethical questions this raises.

Such analysis, as Hayles (2022: 26) argues, 'is only the beginning step in a long process of coming to terms with an uncertain future', yet it exposes the urgent need for approaches and ethical frameworks that are not anthropocentric but which emphasise distributed cognition and decision-making. While this may reveal a complex entanglement of intra-acting components produced through affective flows, always in flux, this also raises ethical reciprocal obligations for us all (Haraway 2016). Acknowledging this complex entanglement does not mean relinquishing our own responsibilities; on the contrary, *all* our actions—co-implicated with other human, nonhuman, material and abstract entities—matter (Barad 2007). Change to deeply integrate ethics into AI education in a transdisciplinary manner cannot come *only* from individual entities—be it teachers, students, disciplines, initiatives or curriculums—rather, all more-than-human components (including those identified in this project) are *mutually* responsible in continually reworking boundaries. It is hoped that, through the experimental methodological framework this project employs, creative and productive engagement with these complex and ethical issues can be encouraged in the context of AI educational design and practice.

Appendices

A1. Participant information and informed consent form

NB: Approved ethics application confirmation paperwork from both University of Edinburgh, and institution from which participants were recruited, has been uploaded to the ethics application dropbox.

Participant information and informed consent form shown overleaf.

Participation in research project:

Cutting together-apart entangled conceptualisations of (machine) 'learning' and 'ethics': assemblage of a 'bricolage-pentimento' artefact through an 'ethico-onto-epistem-ological' approach (Michael Wolfindale)

Section 1: Participant Information

You are being invited to participate in a research study titled *Cutting together-apart entangled conceptualisations of (machine) 'learning' and 'ethics': assemblage of a 'bricolage-pentimento' artefact through an 'ethico-onto-epistem-ological' approach*. This study is being done by Michael Wolfindale from TU Delft, as part of a master's thesis for the MSc in Digital Education at the University of Edinburgh. Before you decide to take part, it is important you understand why the research is being conducted and what it will involve. Please take time to read the following information carefully.

What is the purpose of this study, and why have I been invited to take part?

The purpose of this research study is to explore machine learning students' conceptualisations of (machine) 'learning' and 'ethics', and you have been invited to take part as you are studying a course related to machine learning.

Do I have to take part?

No—it is entirely up to you. If you do decide to take part, please keep this Participant Information sheet and complete the Informed Consent form to show that you understand your rights in relation to the research, and that you are happy to participate.

If you do decide to take part, you are still free to withdraw prior to the completion of analysis (expected to be end of April 2023) and for any reason. In this case, all your data and contributions will be permanently removed. Please note down your participant number (which is on the Consent Form) and provide this to the responsible researcher if you seek to withdraw from the study at a later date. Deciding not to take part or withdrawing from the study will not affect your studies in any way.

What will happen if I decide to take part?

The research will consist of a series of short secure end-to-end encrypted online conversations at different times convenient to you during your machine learning course. During these online conversations you will be invited to share your insights on these themes through text, images, drawings, audio/video, code or any other format you wish. These tasks should take a total of approximately 2 hours to complete.

All data collected will be first anonymised and then analysed, and a 'multimodal' summary of these findings (comprising of mixed media and text analysis) will be produced and will appear in the published master's thesis. As your contributions will be anonymised, there will be no direct reference to you in the final thesis, although anonymised parts of your contributions may appear in some form. However, great care will be taken to remove any personally identifiable information from your contributions to mitigate the risk of you being identified either directly or indirectly.

*(Note: the master's thesis, and all included media—which may include parts or reworked versions of your contributions—will be released under a Creative Commons licence (CC BY-NC-ND 4.0). See **Section 2 point #15** for further information.)*

What are the possible benefits of taking part?

By sharing your experiences, you will be helping the responsible researcher (Michael Wolfindale), the University of Edinburgh and TU Delft better understand how themes of 'learning' and 'ethics' are conceptualised by machine learning students, and how this may be affected by the courses they study. This will help to inform the future design of artificial intelligence programmes and courses, providing further visibility of the social, political and historical contexts surrounding models of (machine) learning introduced, and more deeply integrate ethics through interdisciplinary and transdisciplinary approaches. The responsible researcher will gladly answer questions and provide further details about the research; their contact details are below.

Are there any risks associated with taking part?

The risk of a data breach is always possible and therefore precautions will be taken to secure all data collected and protect your confidentiality; these are detailed below under ***Data protection and confidentiality***. There are also a number of risks detailed in **Section 2B** (see below); steps taken to mitigate these risks are details are detailed in that section. The responsible researcher is happy to answer any questions about these risks and the precautions taken; their contact details are below.

What if I want to withdraw from the study?

Agreeing to participate in this project does not oblige you to remain in the study nor have any further obligation to this study. You can withdraw at any time **prior to the completion of analysis**, estimated to be the end of April 2023 (after this time, your anonymised contributions may be integrated into the final thesis). If you wish to withdraw, please inform the responsible researcher (contact details below) and state your participant code (on the top-right of this sheet). The participant code will be used to identify any anonymised notes or data and permanently remove them from the project.

Data protection and confidentiality

Your data will be processed in accordance with Data Protection Law in the Netherlands; please see below for details about data storage. All information collected about you will be kept strictly confidential; any identifiable data will be accessible only to the responsible researcher, and only anonymised data will be shared beyond this (with the thesis supervisor, or published in the final thesis).

As with any online activity the risk of a breach is always possible, however we will take the following precautions to secure all data collected and protect your confidentiality:

- Once this informed consent form is received via TU Delft institution email, a copy will be stored securely in TU Delft storage designed specifically for research purposes (password protected with a password known only to the responsible researcher), and marked as 'critical' (i.e. containing information that enables the identification of an individual). A unique 'participant code' will be stored securely alongside each participant informed consent form. This code will be used only in the case of a participant withdrawing (see below), in order that anonymised data can be removed from the project.
- From this point onwards, communication will occur only online via the open standard Matrix, for secure, decentralised, real-time communication. The Matrix server will be securely hosted in the Netherlands on a Faculty Managed Server provided by the TU Delft ICT department. All communication will be end-to-end encrypted, only participant Matrix log in details and communication data (text and multimodal contributions) will be stored (no IP addresses will be collected) and only the responsible researcher (Michael Wolfindale) will hold the encryption key (no-one else will be able to decrypt the data even in the unlikely circumstance unauthorised access is gained). To interact with each other via the Matrix server, participants will be able to use any Matrix compatible client they choose (full instructions will be given once the informed consent form is received).
- Any contributions from participants will be anonymised (be removing any information which may directly or indirectly identify a participant), and these contributions will be stored alongside anonymous field notes. The participant code will be used for anonymous storage of these field notes, contributions and any other data produced during the analysis stage, and this will be stored separately in the decentralised open source note-taking and document storage system created by the digital cooperative Anytype (data is stored locally on password-protected devices but transferred between devices via Anytype's end-to-end encryption, see further details). As data is encrypted via a secret keychain phrase that only the responsible researcher knows, and no-one (including at Anytype) will be able to decrypt or access this data.
- A combination of these anonymous field notes, multimodal contributions and analysis will appear in the final thesis (this 'artefact' will itself be in multimodal form and will be inspired by techniques such as 'bricolage' and 'pentimento', drawn from art and other fields).
- All raw non-anonymised data will be removed no later than 30 days following the final marking/moderation of the master's thesis. No raw non-anonymised data will be accessible to anyone other than the responsible researcher (Michael Wolfindale).
- Additionally, a "validity check" will be performed once analysis is complete, whereby participants will be shown analysis of their contribution and given a final opportunity to withdraw or object.

What will happen with the results of this study?

After the research study the de-identified information I provide will be used within the published master's thesis – either through direct quotes or inclusion of media in part or full, or referenced in written or multimodal analysis (that is, consisting of mixed media). In the case of media (such as audio/video) where I may be directly or indirectly identified, media will be converted to another format (for example, text transcription) and anonymised. Quotes or key findings will always be made anonymous in any formal outputs.

The master's thesis, and all included media—which may include parts or reworked versions of your contributions—will be released under a Creative Commons licence (CC BY-NC-ND 4.0). See **Section 2 point #15** for further information.

Who can I contact?

If you have any further questions about the study, please contact:

Responsible Researcher: Michael Wolfindale

Email address: m.j.wolfindale@tudelft.nl

If you wish to make a complaint about the study, please contact:

MSc Digital Education Programme, University of Edinburgh: digitaled@ed.ac.uk

In your communication, please provide the study title and detail the nature of your complaint.

You can get this document on tape, in Braille, large print and various computer formats if you ask us—please contact the responsible researcher on the contact details above.

For information about data management and privacy, please see:

- University of Edinburgh privacy notice for research participants
- TU Delft policies

Section 2: Informed Content (including Explicit Consent Points)

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
A: GENERAL AGREEMENT – RESEARCH GOALS, PARTICPANT TASKS AND VOLUNTARY PARTICIPATION		
1. I confirm I am a student, aged 18 or over, studying a course related to machine learning at TU Delft.	☐	☐
2. I have read and understood the Participant Information above, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	☐	☐
3. I understand that my participation is voluntary, that I can refuse to contribute and that I can withdraw from the study at any time **prior to the completion of analysis** (expected to be end of April 2023), without having to give a reason and without my legal rights or studies being affected. In the case of withdrawal, all my data and contributions will be permanently removed.	☐	☐
4. I understand that taking part in the study involves: • Joining a series of short online secure end-to-end encrypted conversations (via the Matrix open communication standard) at different times during my machine learning course. • Sharing your insights/contributions on the themes of 'learning' and 'ethics' through 'multimodal artefacts' (text, images, drawings, audio/video, code or any other format I choose). *(NB: Some media may be transcribed as text in order to anonymise it and remove any information which may directly or indirectly identify participants. This text will be stored in the decentralised open source end-to-end encrypted note-taking and document storage system Anytype. All raw non-anonymised data will be destroyed no later than 30 days following the final marking/moderation of the master's thesis.)*	☐	☐
5. I understand that the data analysis for this research is expected to be complete by the end of April 2023, the estimated end date of the study is expected to 31 August 2023 and marking/moderation of the master's thesis is expected to be completed before 31 December 2023.	☐	☐
6. I agree to take part as a participant in this study.	☐	☐
B: POTENTIAL RISKS OF PARTICIPATING (INCLUDING DATA PROTECTION)		
7. I understand that taking part in the study involves the following risks, which will be mitigated as detailed below: • The risk of your conversations, contributions or personally identifiable information (such as name and contact details) being subject to unauthorised access due to a data breach will be mitigated through secure storage of data, with use of anonymisation and encryption wherever possible (detailed fully in **point #10**). • The risk of you being personally identified – either directly or indirectly (where your identity is implied or deduced from your contributions) – will be mitigated by the removal of any information from your contributions which directly or may indirectly identify you prior to analysis stage. In the case of audio or video, anonymised text transcriptions will be produced. Raw data (with identifying information) will be securely stored separately to anonymised notes and analysis (detailed fully in **point #10**). • The risk of your identity being deduced from anonymous contribution will be mitigated by the specific course research participants are enrolled in not being mentioned in the final thesis; instead only the level (bachelor's/master's) and subject (machine learning) will be mentioned together with the institution (TU Delft, Netherlands).	☐	☐

• The risk of unauthorised data being accessed once the project has ended will be mitigated by the permanent removal of all raw non-anonymised data no later than 30 days following the final marking/moderation of the master's thesis. • The risk of personal, study or other commitments preventing further involvement in the study will be mitigated by withdrawal for any reason and at any time prior to the completion of analysis being made possible. In this case, all data and contributions you have made will be permanently removed. • The risk of any identifiable data being accessed by your teachers at TU Delft will be mitigated by only the responsible researcher (Michael Wolfindale, Education Advisor), and TU Delft ICT server administrators (for the purpose of maintenance or secure backups), having access. The responsible researcher is not involved in teaching at TU Delft, and does not have any authority over grading. • The risk of you being misrepresented, or of information in the final analysis of your contribution inadvertently (and unbeknown to the researcher) implying your identity, will be mitigated by a "validity check" being performed once analysis is complete; this will involve participants being shown analysis of their contribution and given a final opportunity to withdraw or object.		
8. I understand that taking part in the study also involves collecting specific personally identifiable information (PII) (my name, email address and machine learning course I am studying), and associated personally identifiable research data (PIRD) (online conversation data and contributions) with the potential risk of my identity being revealed (if there is a data breach, or if your identity can be deduced by contributions you make to the research). However, each of these risks have been mitigated as detailed in **point #7**.	☐	☐
9. I understand that – while it is not expected that personally identifiable research data (PIRD) considered as sensitive data within GDPR legislation, such as religion, race or political opinions, will be intentionally stored – it is possible that this may be implied through online conversation data or contributions. To mitigate this, any such data which may directly share or indirectly imply such information will be removed during anonymisation of data and prior to analysis. Additionally, a "validity check" will be performed once the data is analysed, to check with participants that they are happy no such information is directly shared or indirectly implied.	☐	☐
10. I understand that the following steps will be taken to minimise the threat of a data breach, and protect my identity in the event of such a breach: • This informed consent form will be stored, together with a code assigned to each participant, on TU Delft institutional project storage designed specifically for research purposes (password protected with a password known only to the responsible researcher). • Raw communication data and participant contributions will be end-to-end encrypted and stored on a separate server. The encryption key will be known only to the responsible researcher. In the unlikely event of unauthorised accessed being gained to this server, data will be unreadable without this encryption key. • Any further notes and analysis will be anonymised (stored by participant code), and stored separately in the decentralised open source note-taking and document storage system Anytype (data is stored locally on password-protected devices but transferred between devices via Anytype's end-to-end encryption, see further details). Access to this data will not reveal participants' identities, since it is anonymised and stored by participant code (and codes are stored elsewhere). In the case of media (such as audio/video) where I may be directly or indirectly identified, media will be converted to another format (for example, text transcription) and anonymised.	☐	☐
11. I understand that personal information collected about me that can identify me, such as my name and email address, will only be accessible by the responsible researcher (Michael Wolfindale) and will be stored on TU Delft institutional project storage designed specifically for research purposes (password protected with a password known only to the responsible researcher). Raw communication data and participant contributions will be end-to-end encrypted and stored on a separate Matrix server. The encryption key will be known only to the responsible researcher.	☐	☐
12. I understand that the (identifiable) personal data I provide will be destroyed no later than 30 days following the final marking/moderation of the master's thesis (estimated to be before the end of December 2022).	☐	☐

C: RESEARCH PUBLICATION, DISSEMINATION AND APPLICATION		
13. I understand that after the research study the de-identified information I provide will be used within the published master's thesis – either through direct quotes or inclusion of media in part or full, or referenced in written or multimodal analysis (that is, consisting of mixed media). In the case of media (such as audio/video) where I may be directly or indirectly identified, media will be converted to another format (for example, text transcription) and anonymised. I understand that the published thesis may be freely available online via the University of Edinburgh website, or physically via the University of Edinburgh library.	☐	☐
14. I agree that my responses, views or other contributions can be quoted anonymously in research outputs. This may include text or any other media I contribute during the online conversations. A "validity check" will be carried out whereby you approve any contributions prior to submission of the thesis.	☐	☐
15. I agree that the master's thesis, and all included media (which may include my contributions, or reworked versions of my contributions), will be released under the Creative Commons licence Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0). This means that the material can be shared (copied and redistributed in any medium or format) under the following terms: • Attribution will be made to the research project (carried out with the University of Edinburgh and TU Delft), with credit given anonymously to 'all voluntary participants' and with the name of the responsible researcher (Michael Wolfindale). • The material may not be used for commercial purposes. • Remixed or transformed derivatives of the material must not be redistributed (this is to protect participants, since distribution of the materials must be subject to the aforementioned "validity check" with participants who will not be identifiable by anyone other than the responsible researcher).	☐	☐

SIGNATURES

Participant

I agree to take part as a participant in this study.

_____	_____	_____
Name of participant [printed]	Signature	Date

Researcher

I, as responsible researcher, have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____	_____	_____
Researcher name [printed]	Signature	Date

Contact details for further information or questions about the research project:

Responsible Researcher: Michael Wolfindale

Email address: m.j.wolfindale@tudelft.nl

A2. Anonymisation of raw data, and security precautions taken to protect participants' confidentiality

NB: Approved ethics application confirmation paperwork from both University of Edinburgh, and institution from which participants were recruited, has been uploaded to the ethics application dropbox.

All data generated was fully anonymised prior to analysis. Some media was converted to an alternate format (such as reproducing photographs as vector-based imagery) if this better ensured anonymisation of the data.

Participants had the ability to withdraw at any point prior to the completion of data analysis, and to remove their data permanently. In addition, a 'validity' check was performed whereby analysis of their data was shared with them—this step was to ensure they were not misrepresented, and that extracts of their work were not used without their permission (with the opportunity to object or withdraw).

The following steps were taken to protect the confidentiality of participants through secure data storage, and robust procedures to facilitate the anonymisation of data and ability for participants to withdraw (and their data be removed).

Communication with students happened only online, via the secure decentralised open standard Matrix. A Matrix "homeserver" was securely hosted on a managed server provided by the educational institution's ICT department (the educational institution where participants were students).

The Matrix protocol allows communication via text and other modes, and allows sharing of files and multimodal artefacts relevant to the research project. All communication was end-to-end encrypted, only participant Matrix log in details and communication data (text and multimodal contributions) were stored (no IP addresses were collected) and only the responsible researcher (Michael Wolfindale) holds the encryption key (no-one else is able to decrypt the data even in the unlikely circumstance unauthorised access is gained). Participants were not asked to share any personally identifiable research data (PIRD) considered as sensitive data within GDPR legislation, such as religion, race or political opinions; if any information was shared which may directly or indirectly imply these characteristics, it was removed at the anonymisation stage (see below).

Encrypted chat conversations were held individually; no participants were given access to other participants' raw data.

Students who wished to become participants made contact with me (Michael Wolfindale) by emailing my institutional email. I asked students to use their institutional email for this purpose. Details of what the research involved, together with informed consent forms were sent to student participants, and returned via email.

Once informed consent forms were received, these were stored in the institutional project data storage (password protected with a password known only to the responsible researcher), and data in this storage was classified as 'critical' (contains information that enables the identification of an individual) – since the forms contained participant contact details. A unique code was stored alongside each participant informed consent form (see below); this code was used for anonymous storage of field notes and other data produced during the analysis stage. These forms were held as a record of their informed consent, but would also be used in the event that a participant wishes to withdraw. In this case, the code would be looked up, used to identify the anonymous participant data and permanently remove it.

Notes and analysis data were anonymised and stored by participant code in the decentralised open source note-taking and document storage system created by the digital cooperative Anytype (see security details). As data shared between devices is end-to-end encrypted via a secret keychain phrase that only I know, no-one (including at Anytype) could decrypt or access this data.

In summary, for each participant, raw data that is identifiable was stored separately from any subsequent analysis:

- informed consent forms (stored in project data storage, password protected with a password known only to the responsible researcher), with a code attached to each participant.

- Any information within the raw data which may directly or indirectly identify participants was removed prior to analysis, so participant contributions were **fully anonymised**.

- Participants were permitted to withdraw for any reason and at any point prior to the completion of analysis.

A3. Semi-structured introductory conversation for multimodal ethnography

The below shows a suggested opening conversation used as part of the multimodal ethnography, detailed in chapter 3, which was adapted during the data generation stage. This provided a springboard for the ongoing conversations and tasks outlined in Figure 3.d.

> As part of my digital education master's dissertation, I am researching into conceptualisations of learning and how this may intersect with machine learning studies and techniques, as well as ethical issues.
>
> This is a very broad topic, with many different viewpoints and perspectives, and I am focusing on a small aspect of it by setting up secure private one-to-one "chats" with several students currently studying machine learning.
>
> Care has been taken to set up these "chats" in a completely separate server from the institution's teaching and learning infrastructure, and using the open source "Matrix" communications protocol. All chats are end-to-end encrypted—only you and I can access them. The only people with access to the server are me and members of IT who help to maintain the server; however, since the chats themselves are end-to-end encrypted, not even IT staff can read them.
>
> Any content in this chat—be it text, images or anything else—and analysis of it, will always be fully anonymised before appearing in the thesis. In addition, before it gets to that stage, you will have the opportunity to view analysis or excerpts and object/withdraw.
>
> So, you can be safe in the knowledge that you can speak freely and that this research project—while we may discuss your studies—will not have any adverse impact on your studies, nor will it influence your teachers. Your studies always come first and participation is completely voluntary so, if you no longer have time, you can withdraw or simply contribute later on if your schedule frees up.
>
> This is not an interview, and there is no 'right answer' to any questions or discussions I may have with you during the study. It is more of a 'creative exploration' into these broad themes of 'learning' and 'ethics', how this may (or may not) intersect with 'machine learning', and what the implications may be.
>
> I may ask you to 'demonstrate' or 'explain' something during our informal conversations—this can take any form, a drawing or visualisation, piece of code, audio/video or anything else you wish. We will call this an 'artefact'. There are no 'right' ways to do this, and you can be as creative as you like! I may also work together with you on artefacts as part of our discussions—in this sense, we can learn and create together.

A4. Template for field notes and analysis

Figure A4.a below shows the field note and analysis template employed (within the Anytype software) for anonymised logging of field notes and analysis of data fragments.

Field note and analysis template

Field note

Participants Unique participant code

Raw data anonymised before adding to field note. Only unique participant code added here (retains anonymisation, but allows data removal in case of withdrawal).

Analysis notes explanation

These blocks are templates for analysis, placed under data fragments (see below).

General analysis notes

General initial thoughts and analysis.

Multimodal social-semiotic analysis notes

Draws on multimodal social-semiotic theory, informed by Gunther Kress (2009).

Questions to ask (Kress 2009: 57-59):

- **'Whose interest and agency is at work here in the making of meaning?'**
- **'What meaning is being made here?'**
- **'How is meaning being made?'**
- **'With what resources, in what social environment?'**
- **'What are the meaning potentials of the resources that have been used?'**
- **'signs-as-metaphors'** (Kress 2009: 59):
  - the *signified*—_____—is represented by _____ (the s*ignifier)*

Diffractive analysis notes

Draws on a diffractive methodology (Haraway 1997), reading data through new materialist/posthuman theories e.g. Karen Barad's (2007) agential realism/intra-action and N. Katherine Hayles' (2017) cognitive assemblage framework.

What new insights/questions are revealed by diffracting through these theories?

Conversations and artefacts

> Anonymised data fragment (text quote, multimodal artefact etc.)

✏️ **General analysis notes**

Notes

🖼️ **Multimodal social-semiotic analysis notes**

Notes

🔍 **Diffractive analysis notes**

Notes

➡️ *Repeat above for each data fragment.*

References

Barad, K. (2007) *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham and London: Duke University Press.

Haraway, D.J. (1997) *Modest_Witness@Second_Millennium.FemaleMan_Meets_OncoMouse: Feminism and Technoscience*. New York: Routledge.

Hayles, N.K. (2017) *Unthought: The Power of the Cognitive Nonconscious*. Chicago and London: University of Chicago Press.

Kress, G. (2009) *Multimodality: A social semiotic approach to contemporary communication*. London and New York: Routledge.

Figure A4.a: Field note and analysis template.

Produced in Anytype software.

I have uploaded to the submission dropbox fully anonymised versions of the field notes and analysis documents, based on the above template, for all seven participants.

A5. Anonymised conversations, artefacts, field notes and analysis

I have uploaded fully anonymised versions of the conversations, multimodal artefacts, field notes and analysis for all seven participants to the submission dropbox. Please refer to the seven files named **Field note and analysis – participant *x*.pdf**, where *x* is the participant number.

A6. Alternative possibilities for presentation and production of the research-assemblage, and use as part of educational activities

While, for the limited scope of this project, the bricolage-pentimento artefact (Figure 4.h) is presented in a static image format, there are various other possibilities for presentation. This might build upon the static image format through further (potentially collaborative) annotations and comments (expanding on the small number of annotations demonstrated for the limited scope of this dissertation). However, future efforts may also explore 'live methods' for 'real-time' investigation and presentation (Back and Puwar 2012), drawing on 'live sociology' (Back 2012). These possibilities were briefly explored, but were deemed outside of scope since they present a number of time-consuming methodological, technical, accessibility and ethical challenges, particularly given the ethical approval procedures (followed for multiple institutions) and "validity check" with participants. Such challenges would need to be addressed, in particular, how to make some media accessible and inclusive to all (for example, to students who are visually impaired). In fact, it should be noted that before implementation of any multimodal activity in education—including something akin to the bricolage-pentimento artefact (in Figure 4.h)—accessibility should be carefully thought out. While, for the limited scope of this project, efforts were made to supply a number of formats for submission (including a PDF with text readable by a screen reader), the experience would be far from perfect for a visually impaired student or teacher and would need to be implemented together with alternative activities and formats such that there are multiple possibilities for engagement and collaboration in an educational setting.

However, with this outlined, the possibilities revealed through this project appeared methodologically interesting, and may be useful for inspiring future research and/or educational activity design, so they are listed below:

- Three-dimensional model of the research-assemblage, optionally with an aspect of 'extended reality' allowing a visitor to explore the assemblage by moving around/through it. This might be realised through specialist free and open source software, such as Blender, or 'web-based' applications such as Mozilla Hubs and Spoke (the latter of which can be used to build 'three-dimensional worlds'). Web-based applications may allow visitors to communicate and leave comments in a 'live' format.

- An animation of the aforementioned three-dimensional model of the research-assemblage, which conveys the always changing and constantly becoming nature of the affective flows with the research-assemblage.

- 'Interactive' annotated mind map board, with 'frames' highlighting different affective flows and points of interest/analysis, and allowing comments (or even additional data/insights)

to be added in a 'live' format. This might be realised through Miro, Mural, Lucidspark or similar.

Multimodal and collaborative approaches, such as those articulated in this project and suggested above, might additionally be employed as part of educational activities with critical discussion. Through such creative methods, students and teachers could be encouraged to work together to critically consider conceptualisations of (machine) learning—both their own, and those they are collaborating with—and how they might be framed ethically. Perhaps their findings could be mapped out similarly to the bricolage-pentimento artefact in Figure 4.h, using one of the possibilities above or employing another multimodal approach. By drawing on new materialist theories employed in this project—such as affective flows within the research-assemblage (Fox and Alldred 2015; 2016), the cognitive assemblage framework and ethical hierarchy (Hayles 2017; 2022) and intra-action (Barad 2007)—students and teachers might be encouraged to rethink how they conceptualise (machine) learning and the entanglement with ethics. As mentioned in section 4.2.4, Hayles' (2022: 15-16) cognitive assemblage framework, which draws upon Floridi and Sanders (2004) to consider moral accountability and responsibility of agents, might be a starting point—perhaps mapping potential accountability/responsibility to agents multimodally, in something akin to the bricolage-pentimento artefact in Figure 4.h. However, employing a diffractive methodology could subsequently bring it together with other theories for critical discussion which reveals fresh insights and further discussion.

A7. Intra-acting components in the research-assemblage identified from diffractive analysis

A: Component *While shown separated here, components form a flow of affective relations, discussed further below.*	**B: Categories of components of research-assemblage (Fox and Alldred 2015)** ***events to be researched**;* ***participant(s)**;* ***researcher(s)**;* ***research instruments/apparatus**;* ***recording and analysis technologies**;* ***contextual elements***	**C: Hayles' (2017) cognitive assemblage framework** ***cognizers** (actors);* ***noncognizers** (agents), including cognitive supports*	**D: Hayles' (2022) hierarchy of attributes of ethical actors, drawing on Floridi and Sanders (2004)** *ethical actors must possess attributes of the hierarchy:* ***interactivity**—capacity to act as part of a system, with 'intentions that may affect others, creating ethical effects';* ***autonomy**—'ability to make self-directed choices and thus to become morally accountable for them';* ***adaptability**—'capacity to change based on feedback from the environment and from previous choices'* (Hayles 2022: 4)
conceptualisations of learning	events to be researched	mixed—formed through intra-activity between other components	
intersection of conceptualisations of learning with ethics	events to be researched	mixed—formed through intra-activity between other components	
participants	participant	cognizers/actors	(human) ethical actor with degrees of capacity for interactivity, autonomy and adaptability
expectations of participants as 'students', 'professionals', 'coders' etc.	contextual elements	mixed—wider context may involve cognizers/actors acting as ethical actors with degrees of capacity for interactivity, autonomy and adaptability	
pressures on students—financial, academic, careers and employment	contextual elements	mixed—wider context may involve cognizers/actors acting as ethical actors with degrees of capacity for interactivity, autonomy and adaptability	
Generative AI technologies, including ChatGPT and DALL·E	contextual elements (for the purposes of this project, although several experimental conceptualisations of learning generated by these technologies are included in the later bricolage-pentimento artefact)	cognizer/actor	arguably an ethical actor with some degree of capacity for interactivity, autonomy and adaptability
researcher	researcher(s)	cognizer/actor	(human) ethical actor with degrees of capacity for interactivity, autonomy and adaptability
researcher's experiences with machine learning—including previous work with teachers, students, curricula, educational materials, educational technologies, programming languages, research articles and books, etc.	contextual elements	mixed—wider context may involve cognizers/actors acting as ethical actors with degrees of capacity for interactivity, autonomy and adaptability	
theories and methodologies informing this research project—including texts read by researcher	contextual elements	noncognizers/agents	agents within interactive systems (with relations to those producing/employing theories), but with no moral accountability/responsibility in and of themselves
modes, technologies and processes involved in participant conversations, including open source communities involved in these	research instruments/apparatus	mixed—wider context may involve cognizers/actors acting as ethical actors with degrees of capacity for interactivity, autonomy and adaptability	
ethical approval processes and committees	contextual elements	mixed—committees include cognizers/actors acting as ethical actors with degrees of capacity for interactivity, autonomy and adaptability	
modes and technologies involved in presenting the 'cut' of the research-assemblage	recording and analysis technologies	noncognizers/agents	agents within interactive systems (with relations to those producing/employing technologies with specific affordances), but with no moral accountability/responsibility in and of themselves

A: Component *While shown separated here, components form a flow of affective relations, discussed further below.*	B: Categories of components of research-assemblage (Fox and Alldred 2015) ***events to be researched**;* ***participant(s)**;* ***researcher(s)**;* ***research instruments/apparatus**;* ***recording and analysis technologies**;* ***contextual elements***	C: Hayles' (2017) cognitive assemblage framework ***cognizers** (actors);* ***noncognizers** (agents), including cognitive supports*	D: Hayles' (2022) hierarchy of attributes of ethical actors, drawing on Floridi and Sanders (2004) *ethical actors must possess attributes of the hierarchy:* ***interactivity**—capacity to act as part of a system, with 'intentions that may affect others, creating ethical effects';* ***autonomy**—'ability to make self-directed choices and thus to become morally accountable for them';* ***adaptability**—'capacity to change based on feedback from the environment and from previous choices'* (Hayles 2022: 4)
technologies involved in drafting analysis	recording and analysis technologies	noncognizers/agents—although debatable, since some technologies include ChatGPT 'helper' (acting as cognitive support)	agents within interactive systems (with relations to those producing/employing technologies with specific affordances), but with no moral accountability/responsibility in and of themselves (although AI technologies complicate this)
machine learning techniques, the histories of learning theories informing them, (distributed) 'coding practices and pedagogical expositions of machine learning' (Mackenzie 2017: 22) and communities surrounding these aspects	contextual elements	mixed—communities include cognizers/actors acting as ethical actors with degrees of capacity for interactivity, autonomy and adaptability	
teachers	contextual elements	cognizers/actors	(human) ethical actor with degrees of capacity for interactivity, autonomy and adaptability
the course/teaching materials, methods and curricula	contextual elements	noncognizers/agents	agents within interactive systems (with relations to those producing/employing materials, methods and curricula), but with no moral accountability/responsibility in and of themselves
YouTube videos on machine learning	contextual elements	noncognizers/agents—although search/recommender involves algorithms (acting as cognitive support)	technologies involve 'artificial intelligence'/algorithms for search/recommendation
participants' histories with machine learning, and code participants have read and written—including the 'design patterns' that inform this	contextual elements	noncognizers/agents (although debatable, since some programming technologies may involve AI to some degree)	agents within interactive systems (with relations to those producing/employing technologies with specific affordances), but with no moral accountability/responsibility in and of themselves (although AI technologies complicate this)
(open source) coding communities—including its social and cultural norms	contextual elements	mixed—communities include cognizers/actors with degrees of capacity for interactivity, autonomy and adaptability	
the programming languages and frameworks, and surrounding communities	contextual elements	mixed—communities include cognizers/actors with degrees of capacity for interactivity, autonomy and adaptability	
the code involved in 'utilising' or 'applying' machine learning techniques	contextual elements	noncognizers/agents—although programming technologies may involve AI (acting as cognitive supports)	agents within interactive systems (with relations to those producing/employing technologies with specific affordances), but with no moral accountability/responsibility in and of themselves (although AI technologies complicate this)
the modes/semiotic materials/resources used in course materials during their studies	contextual elements	noncognizers/agents—cognitive supports	agents within interactive systems (with relations to those producing/employing resources), but with no moral accountability/responsibility in and of themselves (although AI technologies complicate this)
fiction touching on artificial intelligence—films, books etc.	contextual elements	noncognizers/agents	possibility of technologies involved in script-writing, pre- and post-film production etc. involving 'artificial intelligence'/algorithms to some degree
participants' friends and family	contextual elements	cognizers/actors	(human) ethical actors with degrees of capacity for interactivity, autonomy and adaptability

Table A7.a: Full list of intra-acting components in the research-assemblage from initial diffractive analysis. Summarised in Table 4.a.

A8. Affective flows identified from diffractive analysis

A: Flow *While flows are named for clarity of presentation, all are entangled with one another in the wider research-assemblage—none have privilege over another*	B: Connecting (in no particular order) *What affective flows are produced by the assemblage of relations between human, nonhuman and material entities?*
'Computational' flow	student — students' experiences with computers, databases, processing of data etc. — machine learning applications experienced and developed by students — flowchart and logic diagrams — physics textbook — computational metaphor (brain = computer) — spam email detection example — black box metaphor — assumptions in construction of data models (see Hayles 2022)
'Mathematical' flow	student — squared graph paper — mathematics — Markov — Markov blanket — idea/image of bouncing molecules — Bayesian models and networks — reinforcement learning — agent/environment dualism — maximisation — optimisation — statistics — 'garbage in, garbage out' aphorism — military mathematicians' research into 'electronic brains' (The Hammond Times 1957: 65)
'Behaviourist' flow	student — behaviourist theories of learning — Pavlov's experiments with dogs — punishment — reward — control — reinforcement learning — competition — sport — evolution (survival of 'agents' in 'environment') — agent/environment dualism
'Personal experiences' flow	student — friends studying ethics/neuroscience/psychology — family members' AI textbooks — family members' conceptualisations of learning — personal experiences of learning — lecture from municipality on future engineers and AI
'(Institutional) education' flow	student — teachers — educational programmes — educational materials and activities — educational philosophies — (memories/ideas of) high school physics textbooks — institutional structures — educational initiatives — educational advisors and programme administrators — teaching methods and traditions
'Philosophical perspectives' flow	student — ontology/epistemology — beliefs about 'truth' — technological determinism — notion of progress — technological solutionism — instrumentalism — beliefs about (non-)neutrality of technology — 'garbage in, garbage out' expression
'Ethics' flow	student — beliefs about 'ethics' — beliefs about (non-)neutrality of technology — AI ethics courses — friends studying ethics — notions of control, responsibility, accountability, autonomy, transparency etc. — beliefs about 'bias' — 'garbage in, garbage out' aphorism — ethical approval processes and committees — data security of technologies employed for research
'Media' flow	student — YouTube videos — movies — articles — magazines — hype in popular culture — ChatGPT internet articles — notion of progress
'Research project' flow	student — researcher — theories and methodologies — ethico-onto-epistem-ological approach — technologies and processes involved in participant conversations and artefact production, drafting of analysis and presentation of the 'bricolage-pentimento' artefact — ethical approval processes and committees

Table A8.a: Full list of affective flows from initial diffractive analysis. Summarised in Table 4.c.

References

Albin-Clark, J., Latto, L., Hawxwell, L. and Ovington, J. (2021) 'Becoming-with response-ability: How does diffracting posthuman ontologies with multi-modal sensory ethnography spark a multiplying femifesta/manifesta of noticing, attentiveness and doings in relation to mundane politics and more-than-human pedagogies of response-ability?', *entanglements*, 4(2), pp. 21–30.

Back, L. (2012) 'Live Sociology: Social Research and its Futures', *The Sociological Review*, 60(1_suppl), pp. 18–39.

Back, L. and Puwar, N. (2012) 'A manifesto for live methods: Provocations and capacities', *The Sociological Review*, 60(1_suppl), pp. 6–17.

Barad, K. (1996) 'Meeting the universe halfway: realism and social constructivism without contradiction', in J. Nelson and L.H. Nelson (eds) *Feminism, science, and the philosophy of science*. Dordrecht: Kluwer, pp. 161–194.

Barad, K. (2003) 'Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter', *Signs: Journal of Women in Culture and Society*, 28(3), pp. 801–831.

Barad, K. (2007) *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham and London: Duke University Press.

Barad, K. (2013) 'Ma(r)king Time: Material Entanglements and Re-memberings: Cutting Together-Apart', in P.R. Carlile, D. Nicolini, A. Langley, and H. Tsoukas (eds) *How Matter Matters: Objects, Artifacts, and Materiality in Organization Studies*. Oxford, UK: OUP Oxford, pp. 16–31.

Barad, K. (2014) 'Diffracting Diffraction: Cutting Together-Apart', *Parallax*, 20(3), pp. 168–187.

Baria, A.T. and Cross, K. (2021) 'The brain is a computer is a brain: neuroscience's internal debate and the social significance of the Computational Metaphor', *arXiv [cs.CY]*. Available at: https://doi.org/10.48550/arXiv.2107.14042 (Accessed: 10 March 2023).

Bauman, Z. (1988) 'The second disenchantment', *Theory, Culture & Society*, 5(4), pp. 738–743.

Bayne, S. (2008) 'Higher education as a visual practice: seeing through the virtual learning environment', *Teaching in Higher Education*, 13(4), pp. 395–410.

Bayne, S., Evans, P., Ewins, R., Knox, J., Lamb, J., Macleod, H., O'Shea, C., Ross, J., Sheail, P. and Sinclair, C. (2020) *The Manifesto for Teaching Online*. Cambridge, Massachussetts and London, England: The MIT Press.

Beer, D. (2017) "The social power of algorithms," *Information, Communication and Society*, 20(1), pp. 1–13.

Bøhling, F. (2015) 'The Field Note Assemblage: Researching the Bodily-Affective Dimensions of Drinking and Dancing Ethnographically', in B.T. Knudsen and C. Stage (eds) *Affective Methodologies: Developing Cultural Research Strategies for the Study of Affect*. London: Palgrave Macmillan UK, pp. 161–180.

Borgmann, A. (2009) *Technology and the Character of Contemporary Life: A Philosophical Inquiry*. Chicago, IL: University of Chicago Press.

boyd, d. (2008) “A response to Christine Hine,” *Internet Inquiry: conversations about method*, Thousand Oaks: Sage Publications [Preprint]. Available at: https://www.danah.org/papers/EthnoBoundaries.pdf.

Bozalek, V. and Zembylas, M. (2018) ‘Practicing Reflection or Diffraction? Implications for Research Methodologies in Education’, in R. Braidotti, V. Bozalek, T. Shefer, and M. Zembylas (eds) *Socially Just Pedagogies: Posthumanist, Feminist and Materialist Perspectives in Higher Education*. London and New York: Bloomsbury Publishing, pp. 47–62.

Braidotti, R. (2011) *Nomadic theory: The portable Rosi Braidotti*. New York and Chichester, West Sussex: Columbia University Press.

Braidotti, R. (2013) *The Posthuman*. Cambridge: Polity Press.

Butler, J. (1990) *Gender trouble: feminism and the subversion of identity*. New York: Routledge.

Carlyle, D. (2020) ‘Multiple me, the unfolding ethnographer: Multiple becomings and entanglements are a more-than-human ethnographer’, *entanglements*, 3(1), pp. 31–42.

Cernadas, E. and Fernandez-Delgado, M. (2021) ‘Embedded ethics to teach machine learning courses: an experience’, in *2021 XI International Conference on Virtual Campus (JICV)*. *2021 XI International Conference on Virtual Campus (JICV)*, IEEE, pp. 1–4.

Chen, L., Chen, P. and Lin, Z. (2020a) ‘Artificial intelligence in education: A review’, *IEEE Access: Practical Innovations, Open Solutions*, 8, pp. 75264–75278.

Chen, X., Xie, H., Zou, D. and Hwang, G.-J. (2020b) ‘Application and theory gaps during the rise of Artificial Intelligence in Education’, *Computers and Education: Artificial Intelligence*, 1(100002), p. 100002.

Cobb, M. (2020). *The Idea of the Brain: The Past and Future of Neuroscience.* New York, NY: Basic Books.

Creswell, J.W. and Poth, C.N. (2018) *Qualitative inquiry and research design: Choosing among five approaches*. 4th ed. Christchurch, New Zealand: Sage Publications.

Deleuze, G. and Guattari, F. (1988) *A Thousand Plateaus: Capitalism and Schizophrenia*. Minneapolis, MN and London: University of Minnesota Press.

Dicks, H. (2016) ‘The Philosophy of Biomimicry’, *Philosophy & Technology*, 29(3), pp. 223–243.

Dijkstra, E. (1985) ‘On anthropomorphism in science’. *Philosophers’ Lunch, Austin, TX*, 25 September 1985. Available at: https://www.cs.utexas.edu/users/EWD/ewd09xx/EWD936.PDF (Accessed: 29 May 2023).

Dolphijn, R. and van der Tuin, I. (2012) *New Materialism: Interviews & Cartographies*. Open Humanities Press.

Flewitt, R. (2011) ‘Bringing ethnography to a multimodal investigation of early literacy in a digital age’, *Qualitative Research*, 11(3), pp. 293–310.

Floreano, D. and Mattiussi, C. (2008) *Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies*. MIT Press.

Floridi, L. and Sanders, J.W. (2004) ‘On the morality of artificial agents’, *Minds and Machines*, 14(3), pp. 349–379.

Fox, N.J. and Alldred, P. (2015) 'New materialist social inquiry: designs, methods and the research-assemblage', *International Journal of Social Research Methodology*, 18(4), pp. 399–414.

Fox, N.J. and Alldred, P. (2016) *Sociology and the New Materialism: Theory, Research, Action*. SAGE.

Friesen, N. (2018) *Personalized learning technology and the new behaviorism: Beyond freedom and dignity*. Available at: https://www.normfriesen.info/files/NewbehaviorismFriesen.pdf (Accessed: 7 March 2023).

Geiger, R.S., Yu, K., Yang, Y., Dai, M., Qiu, J., Tang, R. and Huang, J. (2020) 'Garbage in, garbage out?: do machine learning application papers in social computing report where human-labeled training data comes from?', in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. *FAT* '20: Conference on Fairness, Accountability, and Transparency*, New York, NY, USA: ACM. Available at: https://doi.org/10.1145/3351095.3372862.

Grosz, B.J., Grant, D.G., Vredenburgh, K., Behrends, J., Hu, L., Simmons, A. and Waldo, J. (2019) 'Embedded EthiCS: integrating ethics across CS education', *Communications of the ACM*, 62(8), pp. 54–61.

Halliday, M.A.K. (1978) *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.

Halliday, M.A.K. (1984) *An introduction to functional grammar*. London, England: Edward Arnold.

Haraway, D.J. (1988) 'Situated knowledges: The science question in feminism and the privilege of partial perspective', *Feminist Studies: FS*, 14(3), p. 575.

Haraway, D.J. (1997) *Modest_Witness@Second_Millennium.FemaleMan_Meets_OncoMouse: Feminism and Technoscience*. New York: Routledge.

Haraway, D.J. (2016) *Staying with the trouble: Making kin in the Chthulucene*. Durham, NC: Duke University Press.

Hayles, N.K. (2017) *Unthought: The Power of the Cognitive Nonconscious*. Chicago and London: University of Chicago Press.

Hayles, N.K. (2019) 'Can computers create meanings? A cyber/bio/semiotic perspective', *Critical Inquiry*, 46(1), pp. 32–55.

Hayles, N.K. (2022) 'Ethics for Cognitive Assemblages: Who's in Charge Here?', in S. Herbrechter, I. Callus, M. Rossini, M. Grech, M. de Bruin-Molé, and C. John Müller (eds) *Palgrave Handbook of Critical Posthumanism*. Cham: Springer International Publishing, pp. 1–29.

Hennink, M. and Kaiser, B.N. (2022) 'Sample sizes for saturation in qualitative research: A systematic review of empirical tests', *Social Science & Medicine*, 292(114523), p. 114523.

Hickey-Moody, A. (2016) 'A Femifesta for Posthuman Art Education: Visions and Becomings', in C.A. Taylor and C. Hughes (eds) *Posthuman Research Practices in Education*. London: Palgrave Macmillan UK, pp. 258–266.

Hickey-Moody, A. and Willcox, M. (2019) 'Entanglements of Difference as Community Togetherness: Faith, Art and Feminism', *Social Sciences*, 8(9), p. 264.

Hwang, G.-J., Xie, H., Wah, B.W. and Gašević, D. (2020) 'Vision, challenges, roles and research issues of Artificial Intelligence in Education', *Computers and Education: Artificial Intelligence*, 1(100001), p. 100001.

Jackson, A.Y. and Mazzei, L.A. (2012) *Thinking with Theory in Qualitative Research: Viewing Data Across Multiple Perspectives*. London, England: Routledge.

Jacobs, L., Keown, C., Worthley, R. and Ghymn, K.-I. (1991) 'Cross-cultural colour comparisons: Global marketers beware!', *International Marketing Review*, 8(3).

Jacobs, R.A. and Kruschke, J.K. (2011) 'Bayesian learning theory applied to human cognition', *Wiley Interdisciplinary Reviews. Cognitive Science*, 2(1), pp. 8–21.

Knox, J., Williamson, B. and Bayne, S. (2020) "Machine behaviourism: future visions of 'learnification' and 'datafication' across humans and digital technologies," *Learning, Media and Technology*, 45(1), pp. 31–45.

Kress, G. (2009) *Multimodality: A social semiotic approach to contemporary communication*. London and New York: Routledge.

Kress, G. (2011) '"Partnerships in research": multimodality and ethnography', *Qualitative Research*, 11(3), pp. 239–260.

Kuby, C.R. and Christ, R.C. (2018) 'Productive Aporias and Inten(t/s)ionalities of Paradigming: Spacetimematterings in an Introductory Qualitative Research Course', *Qualitative Inquiry*, pp. 293–304.

Lakoff, G. (1992) 'Metaphor and War: The Metaphor System Used to Justify War in the Gulf', in M. Pütz (ed.) *Thirty Years of Linguistic Evolution*. Amsterdam: John Benjamin, pp. 463–481.

Lakoff, G. and Johnson, M. (1980) *Metaphors we live by*. Chicago and London: The University of Chicago Press.

Langley, P. (2011) 'The changing science of machine learning', *Machine Learning*, 82(3), pp. 275–279.

Lasczik Cutcher, A. (2018) 'Pentimento: An Ethnic Identity Revealed, Concealed, Revealed', in Knight, L. and Lasczik Cutcher, A. (eds) *Arts-Research-Education: Connections and Directions*. Cham: Springer International Publishing, pp. 87–100.

Lather, P. and St. Pierre, E.A. (2013) 'Post-qualitative research', *International Journal of Qualitative Studies in Education*: QSE, 26(6), pp. 629–633.

Laupichler, M.C., Aster, A., Schirch, J. and Raupach, T. (2022) 'Artificial intelligence literacy in higher and adult education: A scoping literature review', *Computers and Education: Artificial Intelligence*, 3(100101), p. 100101.

Lim, J.H., Seo, J.E., Choi, J.H. and Kwon, H.Y. (2023) 'A study on the role of higher education and human resources development for data driven AI society', in R. Lee (ed.) *Studies in Computational Intelligence*. Cham: Springer International Publishing, pp. 107–118.

MacCormac, E.R. (1984) 'Men and machines', *Technology in Society*, 6(3), pp. 207–216.

Mackenzie, A. (2017) *Machine Learners: Archaeology of a Data Practice*. Cambridge, Massachusetts and London, England: MIT Press.

Marsh, J. (2017) 'The Internet of Toys: A Posthuman and Multimodal Analysis of Connected Play', *Teachers College Record*, 119(12), pp. 1–32.

Mazzei, L.A. (2014) 'Beyond an Easy Sense: A Diffractive Analysis', *Qualitative Inquiry: QI*, 20(6), pp. 742–746.

Newfield, D. (2018) 'Thebuwa and a Pedagogy of Social Justice: Diffracting Multimodality Through Posthumanism', in R. Braidotti, V. Bozalek, T. Shefer, and M. Zembylas (eds) *Socially Just Pedagogies: Posthumanist, Feminist and Materialist Perspectives in Higher Education*. London and New York: Bloomsbury Publishing, pp. 209–225.

Nwadiugwu, M.C. (2020) 'Neural networks, artificial intelligence and the computational brain', *arXiv [q-bio.NC]*. Available at: http://arxiv.org/abs/2101.08635 (Accessed: 26 May 2023).

Oxford English Dictionary [OED] (2022), *OED Online, Oxford University Press*. Available at: https://www.oed.com/ (Accessed: 14 April 2022).

Papageorgi, I. (2021) 'Positive and negative reinforcement and punishment', in *Encyclopedia of Evolutionary Psychological Science*. Cham: Springer International Publishing, pp. 6079–6081.

Pavlov, I.P. (1902) *The work of the digestive glands*. London: Charles Griffin.

Pink, S. (2011) 'Multimodality, multisensoriality and ethnographic knowing: social semiotics and the phenomenology of perception', *Qualitative Research*, 11(3), pp. 261–276.

Pink, S. (2015) *Doing Sensory Ethnography*. 2nd ed. London: SAGE.

Pink, S. (2021) *Doing Visual Ethnography*. 4th ed. London: SAGE.

Pirini, J., Matelau-Doherty, T. and Norris, S. (2018) 'Multimodal Analysis', in *The Palgrave Handbook of Applied Linguistics Research Methodology*. London: Palgrave Macmillan UK, pp. 639–658.

Saltz, J., Skirpan, M., Fiesler, C., Gorelick, M., Yeh, T., Heckman, R., Dewar, N. and Beard, N. (2019) 'Integrating Ethics within Machine Learning Courses', *ACM Transactions on Computing Education*, 19(4), pp. 1–26.

Sidebottom, K. (2019) *Nomadic enquiry: Reimagining research ethics for 'posthuman' times*, *British Educational Research Association (BERA)*. Available at: https://www.bera.ac.uk/blog/nomadic-enquiry-reimagining-research-ethics-for-posthuman-times (Accessed: 17 May 2023).

Skinner, B.F. (1948) '"Superstition" in the pigeon', *Journal of Experimental Psychology*, 38(2), pp. 168–172.

Skinner, B.F. (1976 [1974]) *About Behaviorism*. Vintage.

Sutton, R.S. and Barto, A.G. (2018) *Reinforcement Learning: An Introduction*. 2nd ed. Cambridge: MIT Press.

Taylor, C. and Dewsbury, B.M. (2018) 'On the problem and promise of metaphor use in science and science communication', *Journal of Microbiology & Biology Education: JMBE*, 19(1).

Taylor, C.A. (2018) 'Each Intra-Action Matters: Towards a Posthuman Ethics for Enlarging Response-ability in Higher Education Pedagogic Practice-ings', in R. Braidotti, V. Bozalek, T. Shefer, and M. Zembylas (eds) *Socially Just Pedagogies: Posthumanist, Feminist and Materialist Perspectives in Higher Education*. London and New York: Bloomsbury Publishing, pp. 81–96.

The Hammond Times (1957) 'Work with new electronic "brains" opens field for army math experts'. Available at: https://www.newspapers.com/clip/50687334/the-times/ (Accessed: 10 April 2023).

Völker, S. (2019) '"Cutting together/apart" – impulses from Karen Barad's feminist materialism for a relational sociology', in U. Kissmann and J. van Loon (eds) *Discussing New Materialism*. Wiesbaden: Springer Fachmedien Wiesbaden, pp. 87–106.

Wang, C., Cai, J., Gao, C. and Ye, X. (2023) 'History, status, and development of AI-based learning science', *SN Computer Science*, 4(3).

Warfield, K. (2016) 'Making the cut: An agential realist examination of selfies and touch', *Social Media + Society*, 2(2), pp. 1–10.

Weerts, H.J.P. and Pechenizkiy, M. (2022) 'Teaching Responsible Machine Learning to Engineers', in K.M. Kinnaird, P. Steinbach, and O. Guhr (eds) *Proceedings of the Second Teaching Machine Learning and Artificial Intelligence Workshop*. PMLR (Proceedings of Machine Learning Research), pp. 40–45. Available at: https://proceedings.mlr.press/v170/weerts22a/weerts22a.pdf