ISSN 2633 1640

Volume 44(2) Winter 2023

Neil McBride
De Montfort University, UK

A Novel Systems Approach to Responsible AI Ecosystems

SYSTEMIST

Publication of
The UK Systems Society

Published by the UK Systems Society

Registered office: Sidelands, Nutgrove Lane, Chew Magna, BRISTOL, BS40 8PU

Registered Charity, No: 1078782

Editor-in-Chief: *Systemist*
Professor Frank Stowell
University of Portsmouth
Portsmouth
Hampshire PO1 2EG
Email: editor-in-chief@systemist.org.uk

Managing Editor: *Systemist*
Dr Christine Welch
Gatcombe House, Copnor Road
Portsmouth PO3 5EJ
Tel: +44 2392 16 0254
Email: editor@systemist.org.uk

All material should be submitted electronically, following the instructions on the Journal website systemist.org.uk. In case of any difficulty, please email the Managing Editor in the first instance. Material should be in word processed form, e.g., Microsoft Word .doc or docx. The font size must be 12 and in Times New Roman, with all figures and tables in a format that will be still legible if reduced by up to 50%. Please supply separate files for figures, in .jpeg or .tiff format.

All materials must conform with the Harvard Referencing System. DOI must be included for all references where possible. A title page must be provided and should include the title of the paper, authors name(s), affiliation, address, and an abstract of 100-150 words. Material published in *Systemist* does not necessarily reflect the views of the UKSS Management Committee or the editors.

A Novel Systems Approach to Responsible AI Ecosystems

Abstract

Artificial Intelligence technology and applications exist within complex digital ecosystems. The development of responsible approaches to the deployment and use of AI requires an approach to ethics which addresses the systemic nature of the complex interactions. This paper attempts to identify a systems approach suitable for analysing and describing AI ecosystems with the aim of identifying interventions and enabling ethical analysis and recommendations. Three classic systems thinking approaches (soft systems thinking, systems dynamics and the viable system model) are considered for their potential to support the development of a system ethics of AI. While all three offer some value, they also exhibit limitations in addressing an AI ethics ecosystem. Attention turns to the Deleuze and Guattari concept of the rhizome to provide a scaffold on which to build a conceptual approach for understanding AI Ethics as an ecosystem. A distinctive approach is described which involves the analysis of assemblages, territories and events. The potential of this approach is illustrated through a case study which analyses the response of three groups (the Future of Life Institute, Centre for AI safety and the British Computer Society) to the release of ChatGPT. The case study illustrates the importance of events as drivers of ethical refection and action, the centrality of assemblages as systems of networks which house worldviews and produce actions, and the nature of a territory which assemblages compete to occupy. It is suggested that adopting a Deleuzian systems approach may offer valuable insights for policy makers at state level, analysts within regulatory organisations and non-governmental organisations and charities at local, national and international levels.

Key Words: AI Ethics, Digital Ecosystems, Systems Thinking, ChatGPT4, Deleuzian Philosophy, Assemblages

Introduction

Technology ethics is a complex field which has been well explored through disciplines including philosophy and sociology. Often the focus is on the properties of the technology and the risks associated with those properties. For example, in

the case of brain computer interfaces, ethics research (TechEthos, 2023) concerned the specific philosophical problems associated with the technology, code and guidelines for its development and use. Ethics is explored in the context of regulation and law concerning, for example the impact of hum rights law on neurotechnologies. Societal impact of the technology is considered in terms of public reaction and perception.

The problem with this approach is that it tends to ring fence the technology, to isolate the technology as the prime source of ethical issues. Our studies of the ethics of technologies airbrush out the plethora of phenomena which impact the technology and set the human environment in which it thrives. Politics, ideologies, beliefs, economic interactions, organisational strategies, national and international agendas, all play a part in the progression of a technologies and the expression of risks and ethical problems and dilemmas. The ethical phenomenon is a systemic phenomenon, expressed through interactions, relationships and the complex organology of technology ethics.

The complex networks of organisations and authorities associated with leading edge technologies; the influence of politics and economics; the cultural and ideological influences; the power structures involving regulator functions as well as influencers and entrepreneurs and thought leaders all point to the systemic nature of technology ethics; to develop approaches which recognised the dynamic, multi-dimensional nature of the technology influence and the ethics.

If we understand technologies, not as isolated artefact with uses to be evaluated, but rather as complex networked phenomena immersed in a sea of politics, economics, social trends and the extensive variety of human behaviour and ideological organisation, we may recognise that we are dealing with complex systems and that there should be a requirement to develop approaches to the exploration of technology ethics which are systems based and give rise to practical systems approach to ethical analysis.

This paper represents one small search for possible system approaches to systems ethics, and develops one, Deleuzian approach, which seems to offer possibilities in understanding the extensive political and power environments within which technology ethics develops. My focus is on AI Ethics, and the understanding of AI Ethics as an ecosystem.

The AI Ecosystem

Recent studies have characterised AI as a social-technical ecosystem where complex interactions take place (Stahl, 2022). This representation is further refined as a responsible AI ecosystem by considering who is answerable for the uses or consequences of the action of the system (Stahl, 2021). While Stahl (2022) calls for a fundamental rethink of AI research, innovation and ethics, treating AI as an ecosystem and identifying interventions, an approach to practical analysis is not offered. The importance of stakeholders and their involvement is recognised, as is the importance of identifying responsibility. Stahl (2022) offers the concept of metaresponsibility, a need to identify responsibility and responsible roles within an AI ecosystem. Identifying those responsibilities within an AI ecosystem may suggest targets for interventions, which may be multiple and diffuse.

The AI ecosystem consists of a wide range of stakeholders operating at a macro and micro level. This will involve major technology companies, governments and international organisations. A wide range of activities may be involved including production of AI, usages, regulation, legislation and development. Key events such as the release of ChatGPT change the landscape in dynamic and possibly unexpected ways. There is a constant evolution and change in the AI ecosystem. Multiple perceptions, ideologies and purposes interact. The ecosystem is characterised by change and instability, driven by the technology and the societal response. Goals, purposes and outcomes are multiple, interacting and in conflict. The is no one underlying purposeful activity that can be ethically isolated.

Given the systems natures of AI, its use in society and the concern over ethics, an approach is needed which, while engaging with the complexity of the environment, provides tools and models which enable a clear portrayal of the systems and the ethical problems and which enable the development of ethical frameworks, policy and guidance. The challenge of a systems approach is to find a way through which offers a holistic lens on what is very much a moving target, with no single worldview nor an easy mapping of the intended transformations.

This paper attempts to identify a systems approach suitable for analysing and describing AI ecosystems with the aim of identifying interventions and enabling ethical analysis and recommendations. Three common approaches, soft systems methodology, systems dynamics and the viable system method, are evaluated for their potential in analysing the AI Ethics ecosystem. I select these because of their popularity and my familiarity with them. Other approaches such as Critical Systems Thinking with its valuable emphasis on boundary critique and

emancipation may be equally or better suited. Indeed, Complex Adaptive Systems Thinking offers a valuable direction, seeking patterns emerging out of complex interaction between agents. Its emphasis on dynamic changes, and sudden shifts, as characterised by chaos theory offer a valuable lens.

Following an evaluation of the three common approaches, a Deleuzian approach to AI Ethics as a system is outlined. Its application is explored using a case study of the response of three organisations, the Future of Life institute, the Center for AI Safety and the British Computer Society to the release of ChatGPT in November 2022. I conclude by identifying the importance of events in driving the system.

Soft Systems Methodology

Soft systems methodology (SSM) may offer the inspiration for understanding stakeholders, relationships, concerns and conflicts in an AI Ethics ecosystem. Rich pictures offer a flexible and creative way of identifying stakeholders and relationships between them. As such, the rich picture provides a good overview of an ecosystem. SSM also offers the possibility of defining the conceptual model in terms of CATWOE and a root definition. Most importantly, SSM offers concepts of the worldview of the system and the transformation which it achieves.

However, AI is a global ecosystem and overlaying an ecosystem of ethics identifies further complexity. As such it exposes some of the limitations of SSM. Firstly, its boundaries are diffuse and hard to characterise. There may be multiple boundaries across which information passes. SSM requires a distinct system boundary which may be hard to identify in systems where the connections are multiple, dynamic and diffuse. The distinction between in the system and out of the system required for SSM becomes difficult to identify in the face of complex interactions. The AI ethics ecosystem is a complexity of patterns and entangled boundaries. SSM requires that a boundary is selected, and the process of selecting a boundary eliminates other options for boundaries may exclude significant elements.

Secondly, the development of a conceptual model requires the selection of a worldview. SSM requires that a system has one worldview which can be stated and underpin the CATWOE model. Ethics ecosystems may harbour multiple worldviews, distinguished by assumptions, ideologies, moral frameworks and learnt experience. Defining one worldview may not work when considering the AI ethics ecosystem. The worldview is a fundamental concept in SSM because it determines the direction of the system, the understanding of the problematical

situation, and the identification of the core conceptual model. In an AI Ethics ecosystem multiple worldviews interact.

Thirdly, in an AI ethics ecosystem, there are multiple transformations, even within a particular worldview. A key assumption of SSM is that the heart of the system is a straightforward "what, how and why" model which, if the dross of social interactions and irrelevancies is washed away, can be defined and optimised for maximum efficiency, effectiveness and efficacy. The complex shifting social dynamics which both hold the systems may mean SSM will be of limited value since in seeking a pure conceptual model it may remove some key complexities.
SSM offers ideas which may be carried forward. The rich picture, which is the familiar representation of a social systems in SSM, is a starting point not an endpoint. The value of the rich picture is the freedom it offers in exploring a system. While often used as a significant output and a vehicle for discussion within an organisation or a situation being investigated, it is actually something to be discarded in the search for the pure essence of the system, the reduction to a conceptual model. Within the investigation step issues of power, culture and intervention purpose are considered in the construction of notes captured as a side task to the developing of understanding of the purposeful system through the rich picture, but may not be pursued or developed.

Key to SSM is the concept of worldview, which remains relevant in understanding an AI Ethics ecosystem, but an AI Ethics ecosystem will contain multiple worldviews. These will be dynamic, influenced by changes in the system, but also anchored in ideologies and social environments. Transformation has ethical consequences, but cannot be pinned down to a single transformation. The concepts of customer and actor are both problematic in that they may be multiple, diffuse and difficult to define.

Systems Dynamics
Systems dynamics focusses on causality, connections, feedback loops and interactions between dynamic variables within the system. As such it provides a holistic view of the multiple interactions in an ecosystem. It is causality driven and seeks to quantify system effects. Systems dynamics origins lie in an effort to compensate for the shortcomings of linear programming. Models that express systems in a constrained number of variables could not accommodate the growing complexity of systems under study. Advances in computer technology allowed the modelling of greater populations of interacting factors, focussing on networks of cause and effect pairings.

Causal loop mapping enables the representation of massive networks of systems variables and their relationships. Hence the effect of point interventions could be explored through computer simulations which are the expected end-products of a systems dynamics study. Further development of system dynamics as popularised in Peter Senge's Fifth discipline sought to characterise systems archetypes, models of common interactions within business systems. As represented in the Beer Game, supply chain vagaries could be easily mapped and understood. Managers could identify simple behavioural patterns in their organisations and implement managerial control mechanisms to counteract faults.

Systems dynamics offered tools for understanding and managing increasingly complex systems within organisation. As such it has had great value in mapping large problems, particularly global problems relating to sustainability. From the early days of Jay Forrester, it has been a tool for confronting global problems of population, food security and climate change. Its potential is still perhaps underutilised as the achievement of sustainable development goals seems less likely to be achieved by 2030.

However, in tackling wicked human and social problems, systems dynamics is essentially a reductionist approach. Systems must be stripped down to measurable variables which have dimensions, represented by values which can go up or down. The representation of the system is a quantisation which is carried further in Senge's systems archetypes. Social landscapes must be represented as numbers; culture, worldview and beliefs discounted unless they can be numerically represented.

The appeal of systems dynamics is reductionist in nature. Uncertainly and impreciseness must be eliminated by identifying variables with clearly delineated dimensions. There should be no room for the type of interpretations that SSM welcomes

In terms of analysing the AI Ethics ecosystem, systems dynamics provides limited insights. While it tackles a wider canvas than SSM, the necessity of quantisation and the primacy of computer simulation is very restricting. Ethics requires debates on uncertainty, conflicting views and social and cultural discussion, all outside its scope. If SSM is prepared to explore written, qualitative narrative which is subject to reflection and interpretation, systems dynamics is not. Perhaps this offers greater appeal to managerial groups who are suspicious of systems philosophies and seek

approaches which, while acknowledging the complexity of human and social systems, still seek the certainty and control of a reductionist approach.

Viable Systems Method

The Viable systems model focusses on understanding organisational structure and developing it beyond simple hierarchical models. Recursion is at the heart of a model which draws its inspiration from cybernetics. VSM recognises the dependencies between organisational functions within an organisation. It counters the danger of reducing organisation complexity too far, citing the need for requisite variety in which the complexity of the organisation matches the complexity of the environment. VSM offers a way of seeing subsystems in the organisation, and characterising the interactions between them. Here a VSM approach may offer insights to the worldview of groups within an organisation which may counter or support an organisational worldview.

Another key aspect is identification of autonomy within an organisation and the autopoietic development of organisation structure. The five systems: Implementation, Control, Intelligence, Co-ordination and Policy provide a good framework for identifying internal ethical issues, expressed both in organisational action and external facing communication through the policy (strategy) system. It also offers an insight into the continuous adaptability of organisations.

Although its hard cybernetic background and the look of its structure suggest a quantitative approach it can be viewed a s soft systems model, with rigor in the softness (Espejo and Gill, 1999). VSM enables the examination of organisational identity through its investigative processes (Lowe et al, 2020). This offers possibilities for identifying and exploring ethical issues within the confines of the organisational boundary. Addressing system 5, policies, may enable AI ethics issues to be described. A comparison with the system 1 activities may identify inconsistencies or even challenges to organisational ethics and policies.

However, the focus of VSM on internal organisation structure, by essentially hierarchical structures, leaves the environment as outside the analysis. In essence VSM pitches the organisation against the environment. An understanding of the AI ecosystem must encompass the environment, multiple organisations and the relationships between them. Therefore, the application of VSM may side-line organisational relationships and reduce the focus to the isolated organisation and its control systems.

A Deleuzian Approach to the AI Ethics ecosystem

It seems to me that none of the major systems thinking approaches reviewed above are sufficient separately for the expression of an AI ecosystem, focussed on the ethics, and the ethical interactions within the system. A system model has to display structure and allow for hierarchies and networks, and recognise that structure is non-reducible, that connections between structures matter as much if not more than the structure or nodes themselves. It has to be multi-organisational, mapping a wide range of organisations and even individuals and enabling understanding of the complex social interactions and the clashes and alignment in philosophies. It must support multiple starting points for tracking (or rather mapping) the system. It must highlight close symbiotic interactions as well as more remote connections, the significance of which may not be immediately obvious. It must recognise the dynamic and uncertain nature of the AI ethics ecosystem, documenting multiple transformations occurring over time and driven by both technological and social events. Additionally, it should encompass the dynamic narrative of the system.

The Deleuze and Guattari concept of the rhizome provides a scaffold on which to build a conceptual approach for understanding AI Ethics as an ecosystem. The rhizome has multiple entry ways, concentration of material give rise to visible expressions in plants. One example is the arbuscular mycorrhizal fungi which form symbiotic links with the roots of plants, providing routes for feedback and nutrient exchange, and connecting plants across species. The activity of mycorrhizal fungo involves not only a colonisation but a connection made between the plant and the soil substrate. The mycorrhizae provides the 'in between': between plant, roots and fungal fruiting body. Extensive mycorrhizal networks connect plants, exchange specific sugars for nutrients generated by the fungi and operate a signalling network (Figueiredo et al, 2021).

The concept of a rhizomic network may provide a way forward in representing an AI ethics ecosystem. It suggests a map where heterogenous connections can be followed and different aspects of the landscape connected and the focus is on the in-betweenness, the connections, the communication of ideas, attitudes and ethical concerns.

In the following paragraphs I identify the core concepts arising from a Deleuzian approach which might be integrated into a framework for considering AI and its ethics as an ecosystem.

Assemblages

Assemblages are a grouping of elements, perhaps ideas, perhaps people or material technology. They are defined as a set of relations which are not permanent, but

emerge. There is a sense of becoming, crystallisation before ultimately dissolving (Wise, 2005). The assemblage is defined by what it can do through the relationships that are formed (Nail, 2017). Delanda describes an assemblages as ' … a symbiosis, a sympathy and depending on alliances and alloys' (DeLanda, 2015). Assemblages can be nested in ways which are reminiscent of recursive elements in VSM and hierarchies within hierarchical theory.

An assemblage may be an abstract representation of an organisation involved in AI, its governance and ethics. We may identify assemblages which act in relationship to adopt a particular responsibility for the actions or outcomes of the AI ecosystem. Assemblages are networks, and themselves are parts of networks of assemblages. While assemblages are only representations, there can be types of assemblages. Abstract assemblages represent concepts which draw in concrete assemblages or are promoted by concrete assemblages. Indeed, a concrete assemblage could be an embodiment of an abstract assemblage.

An assemblage may be characterised by relationships with other assemblages including allies and enemies. As such it is the active and living relationships within the elements of the assemblage and between assemblages that give it its expression. Its social existence is only temporary, however permanent it may seem. Like arches in a building, active forces hold an assemblage in place. Any shift of ground might result in its collapse and fragmentation.
Assemblages are sustained by perceptions of purpose, function and outcome driven by the worldview(s) of those connected with them. Assemblages themselves may be unstable, and form unstable relationships. Although an assemblage is highly abstract, its representation may be crystallised through organisational forms, administrative processes, marketing strategies, and regulations, for example. Assemblages have content and expression; patterns of practice and elements. Therefore, we need to attend to the language of the system as it develops from and expresses the worldview.

Nail (2017) identifies three aspects of an assemblage. Firstly, in the abstract an assemblage consists of a group of relations clustering around an idea, concept, political ambition. It has a particular focus on time and space. The abstract machine lays out the relations which give rise to concrete structure and agency. The abstract assemblage a concept of a mint with a hole, a recipe, a supply chain give rise to concrete elements, the factory, machines, lorries etc. Secondly, the concrete assemblages give embodiment to the abstract. There is reciprocity between the abstract and the concrete assemblage. Consequences and limitations derive from

the concrete assemblage. Thirdly, an assemblage is always a collective of elements from which a persona emerges. It is always relational inside and out. No one is subject to him or herself but to others; it is in that intersubjectivity that ethical issues reside.

An assemblage has a linguistic expression; connections within an ethical ecosystem are by alignments and misalignment of those expressions, by connecting of meanings. Hence part of our investigation of the AI ethical ecosystem involves assessment of collective expression, understanding and meaning. This collectiveness appears as a phenomenon through strategy statements, codes of ethics, regulations and acceptable rules. Assemblages group into large scale networks of constellations, enabling the emergence of cultures; local acts can be abstracted to the global.

Responsibility emerges in between the assemblages, Responsibility is relational: an assemblage is responsible to another assemblage for the outcome of the AI ecosystem. Socially, assemblages are power structures. Emanating out of them are particular worldviews which in terms of social assemblages, members or participants adhere to. The assemblage is held together by general adherence to a worldview and acceptance of power structures. Furthermore, assemblages link with events. Events change assemblages, push assemblages forward.
Finally, it should be noted that an assemblage is more something dynamic, existing in its acting and its agency. The tendency to freeze the concept into fixed structure needs to be avoided. The focus is on agency, an explicit bringing together of the parts. There is an underpinning concept of acting, bringing together in an event. Hence the coming together of element in an assemblage and indeed their disassembling are events, and it is through these events that ethical issues can be identified.

The concept of assemblages offers a way of thinking in which connections between worldview and knowledge systems may be explored (Dewsbury, 2011). They incorporate ideas of networks, flow and emergence (or becomings).

In terms of AI ethics, our main concern will be the social assemblages, concentrations of social relations from which ethical dilemmas and concerns emerge. Primarily social, human assemblages are those concerned with territorialisation. As such they encase power relations. Rather like strong electromagnets can hold a plasma in place, an assemblage of social relations dynamically contains power.

Territoralisation

In concert with the concept of assemblage is that of territoriality, which applies the metaphor of territory to analyse how the assemblages occupy social space. Assemblages occupy and remove themselves from territories. These territories may be "assemblages of enunciation", language structures which express ideas in flux. They can concern ideas and theories, interpretations of phenomena, technologies and technology applications, influenced by worldviews, intentions, focus, expectation and assumption. Territories tend to be more stable, with clearer boundaries. They are more passive structures, although territorialized, de-territorialized and re-territorialized by evolving assemblages. Assemblages operate on territories, they can be seen as processes of putting together, of mapping out a territory.

Territories are socially constructed. They are characterised by distinctive boundaries. A key aspect of a Deleuzian view of territory concerns difference. An exploration of a territory should involve an analysis of difference. Rather than seeking a global normative ethics, a critical focus is on the differences, not only differences in the territories which define boundaries, but differences in assemblages which seek to de-territoralise and re-territoralise. Territory establishes critical distance (Aurora, 2014); which concerns distance between different philosophies, perspectives and goals. By understanding differences in both the territories and the terrritoralisation by assemblages we can, to recruit a concept from SSM, identify *accommodations*: ethical frameworks that conflicting assemblages can live with, if not providing an ideal fit to different moral frameworks and assumptions.

In other words, in a social world where moral arguments may be continuous and irresolvable (MacIntyre, 2007), it may not be possible to establish absolutes and normative guidelines, nor might it be acceptable to retain a relativeness where any moral lens is acceptable. Rather than a fortress ethics in which assemblages fight for possession of ethical territories, a community ethics might be imagined, managed through accommodations which different assemblages can live with.

A transformation of a territory occurs through a process of de-territoralisation and re-territoralisation by assemblages. A transformation may involve control over phenomena, relationships, markets, messages, access to resources and so on. The outcome of the transformation may be new behaviours, expressions, actors, and interpretations. A re-territoralisation may establish regulations, laws and rules.

Hence a territory is a subject of power. The interpretation and development of a territory through de-territorialisation and re-territorialisation is the activity of assemblages which define and contest boundaries, and seek to occupy through the stating of meaning. This occupation of territory is expressed though organisation of the territory, regulation, authentication and validation. Territorialisation involves the delimiting of abstract or concrete space, the exerting of control over meaning, resources or access by whatever assemblage. In ethics this might be regulatory organisations, academia, commerce or state, arguing primacy, intellectual superiority or domination in terms of resources and property. It may involve arguments concerning extent of knowledge and understanding about the territory: "ignorant actors who need to be reorganized and disciplined via new socio-spatial arrangements engineered by the state" (Bassett and Gautier, 2014) It may involve alliances between assemblages and attempts to shift or redefine the territory.

Events

Ethical problems tend to rise to the surface in response to events. Their apparent significance relates to a happening in time and space. In considering assemblages as a basis for exploring the ethics of an AI ecosystem, the questions move from the "what, how and why" of SSM to "when, where and how". The focus is on events and the ethics of technology emanating from them. An event may be underpinned by technology, perhaps the release of a new machine, but it is primarily a human, social event.

The ethical significance of a technology lies in the social events that triggers its use or trigger a concern about its effect. For example, in the case of Apple AirTags, its ethical significance was triggered by media reports of its use to stalk a celebrity. The use of AirTags for tracking people had not been anticipated by developers (McBride, 2024). Similarly, the implementation of the GDPR imposed ethical significance on a range of technologies. Also, the release of ChatGPT created ethical concerns which were not previously apparent.

Events may not be isolated, but may be sequences or families of events. They may involve particular technological advances occurring, but may be equally arise from social change or economic change. The use of a particular technology and therefore the emergence of ethical issues may not depend on the technology, but on social, economic and commercial changes. For example, electric cars are nothing new but events in battery technology, climate change, government policy and manufacturer decisions trigger ethical issues relating to self-driving cars. Events are

differentiating. In a rhizomic system they can emerge from and effect any point in the network. Importantly an event indicates change occurring as well as initiating change. Analysis should address not just large events, features in the media with global reach, but micro events in individual lives or in groups and communities.

Thus, the task of anticipating ethical issues should be to consider events past, present and future; to identify those events in the social milieu and consider the effects of interventions in an AI ecosystem. Surprisingly, an event-driven approach to AI ethics suggests a role for the causal loop mapping at the heart of systems dynamic in mapping cause and effect as the event triggers the cascade of technological ethical issues through the system. Our starting point is not the technology and the apparent characteristics of the technology, but the events which trigger ethical concerns. "[T]hey reshape the conceptual and material fabric of connectivity, relationships, pathways and institutions" (Beck and Glayzon, 2016). As such they are a source of transformation. Events may reconfigure space. In the digital they may affect people disparately located in space.

Significantly, events may emanate from social assemblages. Thus, events triggered by an assemblage can reconfigure the assemblage or even dismantle it. Events reset practice in response to emerging effects and consequences. Hence events can change the ethical focus. Issues and practice which was perhaps not imagined, or ill-addressed in ethical assessments, becomes central and of prime concern. Thus, events may determine the ethical agenda and even trigger ethical concern which was previously abstract. The event may change what an assemblage can do, or where and when it can act. It may reset consequences and outcomes.
Events will change the reality of a situation. They are the source of transformation, points in time and space. They may trigger assemblages to change the territories they occupy. A de-territoralisation and re-territoralisational may result from an event and an as such concretise a transformation.

Case Analysis: AI, Force for good or existential threat?
I will illustrate the potential of a Deleuzian approach using three media events in which groupings of stakeholders embedded in assemblages issued media statements concerning the potential effect of AI on humanity. The starting point of the analysis is the sequence of events which generate ethical discussion and highlight ethical concerns. These events showcase three relational groupings into assemblages, and focus on a public territory which concerns the abstraction of the public perception of AI.

Events are often tethered to previous events and the origin or the three events highlighted in this discussion is in the public launch of CHatGPT on 30 November 2022 by OpenAI. As this large language model caught the public's eye, its power and potential in fields from drugs development to education were touched on. Public users began to realise the potential of such models, particularly in education. In response to the accelerating profile and concern about AI, three assemblages issued prominent public statements aimed at territorialising the public landscape of debate. The following discusses the events, assemblages and territory.

The Events

Events provided the driver for ethical reflection and the development of guidelines. They may be ordered in various ways. Date order provides the most obvious approach. But they could be ordered in terms of impact assessed by a selected measurable.

The first event was an Open Letter published on March 22 by the Future of Life Institute, signed by some 30,000 people including Elon Musk. It was a response directly to the release of ChatGPT4 eight days earlier. It called for a six month pause in large language model training after the release of GPT 4. Their letter argues that "*Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources*" (FLI,2023a). AI is becoming human-competitive. The purpose of the pause would be to generate shared safety protocols which ensure that systems adhering to them are safe beyond a reasonable doubt.

The second event was a statement released by the Centre for AI Safety on 30 May. The one-sentence statement read; "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war" (CAIS,2023). Signatories included Tegmark, Kurzweil, and Gates as well as the CEOs of Deepmind, OpenAI and Anthropic. It received coverage in The New York Times and Time Magazine. Its theme was to identify AI as an existential risk and hence direct attention to the Centre for AI Safety which it offers a structured approach to mitigating AI risks.

The third event was triggered by the British Computer Society (BCS) a UK-based grouping of primarily computer science academics which promotes computer science and offers certifications of computer science courses in universities. Issued on 7 June, and signed by 1370 people, it explicitly counters the two previous statements: "AI is not an existential threat to humanity; it will be a transformative force for good if we get critical decisions about its development and use right" (BCS.2023). It is

directed to the UK Government and highlights the potential role of the UK. IT calls on the UK to lead the way in setting technical and professional standards for AI and hence “"Coded in Britain" can become a global byword for high quality, ethical, inclusive AI.”

Each event is an ethically motivated response to a previous event, an attempt by an assemblage, based on the ideology and activities of that assemblage, to territorialise the landscape of public debate. The event raises ethical concerns, perhaps explicitly because of the nature of the events as is the case here, or as an outcome of a social or technical event. Examining the nature of the assemblage that generated the event and the characteristic of the territory targeted by the assemblage, we can identify specific ethical concerns and understand better the dynamic and systemic nature of the ethical ecosystem.

The Assemblages

Each of the assemblages which initiated an event are sets of relations between elements which are assemblages in themselves, or representatives or figureheads, embodying the spirit of the assemblage.

The Future of Life Institute (FLI) focuses on the large-scale risks of transformative technology, particularly biotechnology and AI and reducing large scale harm and existential risk. Its activities include policy development, education and research support. Its values are underpinned by science and reason and are inclusive: “people are not instrumental, which means that positive impact on the world should be achieved while maintaining integrity, kindness, and respect for others.” (FLI, 2023b). The FLI is a grouping of university computer science academics. While the board of directors includes the founder of Skype and a researcher from DeepMind, it is essentially an assemblage of AI technologists, led by a professor of the physics of information. FLI operates as charity, underpinned by an endowment. The FLI covers a wider scope of existential risks than AI, but had concentrated on AI. FLI raises the spectre of existential destruction, for example, through a fictional presentation of the escalation of nuclear war triggered by AI. However, their “Imagine a World” podcast, looks to identifying plausible and positive futures using AI, seeing AI as having potential to create community.

The development of the list of signatories, while originating with the FLI is itself an assemblage, a set of relations. While many, if not most, signatories are academics. There are some commercial CEOs and researchers in companies, headliners including Elon Musk and the CEOs of Apple, Stability AI and Getty

Images. However, the assemblage expressed by the list is rather an extension of the FLI assemblage in pursuit of a territory.

The Centre for AI Safety (CAIS) is concerned with managing AI risk, reducing societal-scale risk through research, education, policy and what is described as field-building. Field building involves supporting a wider research community through the provision of computing capacity for training large language models, the provision of philosophy fellowships and the support of academic speaker programmes. Furthermore, CAIS intends to build a research ecosystem which promotes safe AI. What they do is driven by: advancing AI safety research, building the safety community, and promoting safety standards. The CAIS is connected with the University of California, Berkeley and directed by Dan Hendrycks, whose research is strongly focus on AI Safety. The worldview of the CAIS, at least represented by its founder, see AI as evolutionarily fitter than humanity and hence likely to occupy the global environmental dominant position which humanity occupies.

CAIS, as a tight community based in San Francisco, developed the one-line statement and an extended assemblage which included CEOs of Google, OpenAI and Anthropic. Geoffrey Hinton and Yoshua Bengio, two of the three researchers who won a Turing Award for their pioneering work on neural networks. There is some overlap with the FLI signatories, but a more commercial emphasis. Significantly, semi-official endorsement of the statement by Google, OpenAI and Anthropic is separate from the FLI open letter. And the endorsement by OpenAI is quite explicit. They see AI creating societal scale disruptions with risks serious enough to warrant government intervention.

The British Computer Society (BCS) is a UK professional organisation. primarily academically based, which offers validation of computer science degrees, and has both members and fellows. Local branches are run from universities as well as companies. As a professional organisation, it seeks to connect industry and education, and has a strong focus on education. The central structure is supported though membership fees. Its activity is in promoting careers in computing. It is a set of professional relationships, centred on a range of university departments. Formed in 1957, it has a specific UK focus. Helping students progress their careers is important, and the BCS highlights IT improving society. The BCS activities more address the present, considering current technology, offering qualifications and certification.

The BCS view of AI concerns regulation of use rather than technology : "attempting to halt AI development would not work; Not every country or organisation would obey, and bad actors would gain an advantage." Although research would be needed to confirm, it is likely that all signatories are members or fellows of the BCS.

These three assemblages are dynamic groupings of relationships, each with a loosely held ideologies, each developing and changing, each founded in academic, university environments. Two are recent aggregations, while BCS has its origins in the foundations of computing over sixty years ago. Each has spawned a new assemblage through the signing of letters. BCS's remains internal while FLI and CAIS extend outwards.

Each assemblage contains many relationships which enable the maintenance of an ideology a philosophical direction, a plasma is held in place and driving the effort of territorialisation.

The Territory

For each assemblage, the producing of an open letter is an act of territorialisation, which simultaneously intends to occupy the territory and define it. The territory is an AI Ethics ecosystem, a mix of technologies, technology companies, governments and regulators of AI. It is a landscape flooded by public and national perceptions of AI and its potential and risks. The nature of the AI ethics ecosystem territory is semiotic, it is a system of meanings. It concerns the definition of AI, the nature of the technology, the potential and actual applications and the prophesies of its future.

As such the territory is shifting, vague, ethereal, a dynamic system of changing ideas about the nature of AI, one which is free-floating. Territorialisation involves the definition of boundaries, moved from the subjective and crystallised in some form. The activities of the assemblages are directed toward territorialisation, through interventions of such a nature that define access, regulation, and the legal environment.

A common goal of territorialization is to govern people and resources located within and around the territory (Scott, 1998). The events are attempts by each assemblage to exert control over the meaning attached to the territory through exposure in the media and books. Thus, the AI ethics ecosystem, targeted by each assemblage, is a set of ideas and meanings which through regulation and laws

exerted by a wide range of stakeholders will be the conceptual substrate which supports the economics, technology and application of AI.

Discussion

This brief case study illustrates the importance of events as drivers of ethical refection and action, the centrality of assemblages as systems of networks which house worldviews and produce actions, and the nature of a territory which assemblages compete to occupy.

Identification and analysis of events provides a reasonable starting point for a Deleuzian analysis of a complex human ecosystem. Crucially, events occur, things happen. These draw technology in, create change through the actions, doings, intervention of assemblages which change, reset territories. Macro, meso and micro events are occurring all the time, creating flows and dynamics which influence territories and evolve a constantly developing ethical environment. In Deleuzian terms there is no beginning or end, no fixed point of departure. In a rhizomic network events, intervention and exits can occur anywhere. Assemblages constantly interact and relate in exploring and contesting territories in response to flows of events. Events have ethical consequences which need exploration. Events attract technology; hence the technology ethics should be interpreted in the light of system events. Systemically, events suggest or trigger connections, establishing and resetting relations, challenging and moving boundaries; crossing boundaries, breaching boundaries, changing permeability, reassigning meaning to signals, enabling new and evolving signals.

Assemblages are complex dynamic entities structured through internal relationships and relationship to the exterior. The evolution of an assemblage is a continuous process and its behaviour, its response to its exteriority is emergent (Hanseth and Modol, 2021). Assemblages both fragment and coalesce in a dynamic fashion. Besides acting on a territory, an assemblage can be part of a territory, subject to territorialisation itself. Also, assemblages can develop in response to events and accumulate actors, elements and relationships. In the case study, the BCS assemblage emerges as a response to an event, while the pre-existing Future of Life and CAIS assemblages respond to the release of CHATGPT3.

AI Ethics is phenomenological, not solely a technology nor an information architecture but a complex human atmosphere and environment where the complex interaction of individuals and societies, commercial concerns and governments, ideologies and powers create a constant dynamic flow of conflict and collaboration

in an immersive system. Any systems approach needs to embrace ideological, political and psychological aspects while providing a means for structured analysis and representation.

Any system model, including soft systems methodology, systems dynamics and the viable systems method may offer valuable insights. Although I critically evaluate three potential classic systems approaches, these offer important insights of their own. Soft Systems offers insight into the landscape, particularly expressed in terms of stakeholder issues. System dynamics supports the event driven nature of this Deleuzian approach by helping to identify causal cascades that might be triggered by events. The value of Viable Systems Method lies in its recursive analysis of organisational structure which has value in the analysis of assemblages. Other systems approaches should not be excluded. Complexity theory resonates strongly with the concept of assemblages. Critical systems thinking, amongst its wise insights, encourages the questioning of boundary judgements, which is an important concern in the analysis of assemblages and territories.

However, in pursuing a Deleuzian approach my intention is to go beyond the addressing of information architectures and flows towards politics, ideologies, psychologies, and human behaviour in the large; to find ways forward to examine ethical environments from a systems point of view, extending the landscape beyond foci such as human rights frameworks and codes of ethics.

Ethics is relational and the AI ethics ecosystem which I am seeking to characterise and explore through a systems approach is built out of human relationships and exchanges, within complex groupings, constantly evolving assemblages, fed by exchanging messages, creating and dissembling meanings and acting in response to those meanings. Technology is necessary but not sufficient for the development of an AI ethics ecosystem. The information architecture enables or carries the architecture of human relationships.

Systems thinking should be underpinned by a practical concern for application. Methodologies such as soft systems and systems dynamics provide practical tools for application in organisation in confronting wicked problems and finding potential interventions. While this study is conceptual in nature, it is important to consider the development of practical frameworks and application. A Deleuzian framework would work on a wider canvas, both beyond organisational concerns and in tracking the wider political and ideological environment. There is a need to develop processes by which an extensive, even global, technologically oriented

system can be explored. Starting with events which trigger ethical ideas, discussion, problems and issues, territories can then be identified and scoped. Models of assemblages could be developed and their changes in form tracked over time. This approach may enable calling out of ethical inadequacies, identifying questions, raising concerns and identifying intervention points in engaging with assemblages. The approach may trigger or extend practical debate and dialogue with the aim of challenging and changing policy within institutions and supporting ethical human behaviour within complex ethical landscapes.
Identifying social groupings on a global scale, teasing out ideologies, agenda and power relations, and understanding the dynamics of occupation of a territory may all be facilitated. Such an approach may have value for policy makers at state level, analysts within regulatory organisations, think tanks and research institutes, as well as non-governmental organisation and charities at local, national and international levels.

Conclusion

In attempting to anticipate ethical consequences of technology we are too technologically focussed. The ethics does not emerge de novo from the technology itself, but rather through the social environment and the events which make it relevant.

The aim of this work has been to pursue a systems view of technology ethics, particularly focuses around the substantial issues of AI ethics. As such, AI Ethics cannot be reduced to a risk analysis based on the characteristic of a particular use case, nor an isolated technology with ethical attributes and issues. The technology cannot stand alone, nor be attributed a life of its own autonomous from the human entanglement.

In this paper I have tried a Deleuzian approach on for size, condensed to three conceptual dimensions: assemblages, territories and events in the hope it might enable a better understanding of AI Ethics, and also illuminate practical approaches and the potential for intervention. Crucially, it is important that we shift our thinking in AI Ethics towards systems thinking.

References

Aurora, S. (2014). Territory and subjectivity: The philosophical nomadism of Deleuze and Canetti. *Minerva,* 18, 1-26.

Bassett, T. J & Gautier, D (2014). Regulation by territorialization: The political ecology of conservation and development territories. *EchoGéo*, 29. URL:

http://journals.openedition.org/echogeo/14038 ; DOI: https://doi.org/10.4000/echogeo.14038

BCS (2023, April 24) https://www.bcs.org/articles-opinion-and-research/bcs-open-letter-calls-for-ai-to-be-recognised-as-force-for-good-not-threat-to-humanity/

Beck, C. & Glayzon, F-X. (2016). Deleuze and the event(s). *Journal for Cultural Research*, 20 4), 329-333. DOI: 10.1080/14797585.2016.1264770.

CAIS (2023, April 24). https://www.safe.ai/statement-on-ai-risk

DeLanda, M. (2015). *Assemblage Theory*. Edinburgh University Press.

Dewsbury, J-D. (2011.) The Deleuze-Guattarian assemblage: Plastic habits. *Area.* 43(2) 148-153.

Espejo, R. & Gill, A. (1999) The Viable Systems Model as a framework for understanding organisations. http://www.moderntimesworkplace.com/good_reading/GRRespSelf/TheViableSystemModel.pdf

Figueiredo AF, Boy J. & Guggenberger G. (2021). Common mycorrhizae network: A review of the theories and mechanisms behind underground interactions. *Front. Fungal Biol.* 2:735299. DOI: 10.3389/ffunb.2021.735299.

FLI. (2023a, April 24). Future of Life Institute Call for pause.
https://futureoflife.org/open-letter/pause-giant-ai-experiments/

FLI. (2023b, April 24) Our mission https://futureoflife.org/our-mission/

Hanseth, O. & Modol, J. R. (2021). The dynamics of architecture-governance configurations: An assemblage theory approach. *Journal of the Association for Information Systems*, 22(1). DOI: 10.17705/1jais.00656.

Lowe,D., Espinosa.A. & Yearworth,M. (2020). Constitutive rules for guiding the use of the viable system model: Reflections on practice. *European Journal of Operational Research*, 287(3), 1014-1035. https://doi.org/10.1016/j.ejor.2020.05.030.

McBride, N. (2024). Apple Airtags as people trackers. In Tracking People. Ed. Hucklesby,A & Holt,R. Routledge.

MacIntyre, A. (2007). *After Virtue*. Duckworth.

Nail,T. (2017). What is an Assemblage? SubStance, 46(1),21-23.

Scott J. (1998) *Seeing like a state: How certain schemes to improve the human condition have failed.* Yale University Press, New Haven, CT.

Stahl, B. (2021.) *AI for a Better Future*. Springer

Stahl, B. (2022). From computer Ethics and the ethics of AI towards an ethics of digital ecosystems. *AI and Ethics,* 2, 65-77. https://doi.org/10.1007/s43681-021-00080-1.

TechEthos (2023) Neurotechnologies https://www.techethos.eu/neurotechnologies/

Wise, J.M. (2005). Assemblage. In Stivale, C.J. (Ed.), *Giles Deleuze, Key Concepts*. Acumen, Chesham.

Acknowledgement

I thank the anonymous referee for very helpful insights

UK SYSTEMS SOCIETY

UKSS Gold Medallists

Professor Russell Ackoff
Professor Stafford Beer
Dr Fritjof Capra
Professor Peter Checkland
Professor C. West Churchman
Professor Humberto Maturana
Sir Geoffrey Vickers

Systemist **is a publication of**
The United Kingdom Systems Society