Journal of Responsible Technology 19 (2024) 100087

Contents lists available at ScienceDirect

Journal of Responsible Technology

journal homepage: www.sciencedirect.com/journal/journal-of-responsible-technology

Infrastructural justice for responsible software engineering☆

Sarah Robinson [a,b,*], Jim Buckley [a,c], Luigina Ciolfi [a,b], Conor Linehan [a,b], Clare McInerney [a,b,c], Bashar Nuseibeh [d], John Twomey [a,b], Irum Rauf [a,b,d], John McCarthy [a,b]

[a] Lero - SFI Research Centre for Software, Ireland
[b] School of Applied Psychology, University College Cork, Ireland
[c] School of Computer Science and Information Systems, University of Limerick, Ireland
[d] Open University, School of Computing and Communications, Walton Hall, Milton Keynes, United Kingdom

ARTICLE INFO

Keywords:
Responsible software engineering
Infrastructure
Social connection model of responsibility
Installed base
Deepfake technology

ABSTRACT

In recent years, we have seen many examples of software products unintentionally causing demonstrable harm. Many guidelines for ethical and responsible computing have been developed in response. Dominant approaches typically attribute liability and blame to individual companies or actors, rather than understanding how the working practices, norms, and cultural understandings in the software industry contribute to such outcomes. In this paper, we propose an understanding of responsibility that is infrastructural, relational, and cultural; thus, providing a foundation to better enable responsible software engineering into the future. Our approach draws on Young's (2006) social connection model of responsibility and Star and Ruhleder's (1994) concept of infrastructure. By bringing these theories together we introduce a concept called infrastructural injustice, which offers a new way for software engineers to consider their opportunities for responsible action with respect to society and the planet. We illustrate the utility of this approach by applying it to an Open-Source software communities' development of Deepfake technology, to find key leverage points of responsibility that are relevant to both Deepfake technology and software engineering more broadly.

1. Introduction

There is a growing concern among academia, industry, and representative bodies with defining and developing ethical and responsible Software Engineering (SE) practices. For example, there are over 170 sets of guidelines related to Artificial Intelligence alone, ranging from company principles and checklists to multilateral guidelines (AlgorithmWatch, 2023; Deshpande & Sharp, 2022). The ACM (Gotterbarn et al., 2018) has a long-established codes of conduct aimed at more general Software Engineering. The code focuses on 8 ethical principles linked to software engineering practice, including that software engineers should only approve software in line with the public interest or, 'if they have a well-founded belief that it is safe, meets specifications, passes appropriate tests, and does not diminish quality of life/privacy or harms the environment.' [(Gotterbarn et al., 2018), p. 112]. The code also states that engineers should disclose any potential dangers to users and the environment associated with software to an appropriate authority. Yet we know that there are staggering environmental costs in the generation of Large Language Models that underpin many software systems (Bender et al., 2021a), and in the cloud infrastructure many such systems rely on (Ensmenger, 2021; Monserrate, 2022), harms that clearly fall into the category of 'hurting others' and 'discrimination' regularly occur; for example, see (Amnesty International, 2022; Bender et al., 2021a; Birhane, 2022; Buolamwini & Gebru, 2018; Charitsis & Lehtiniemi, 2023; Costanza-Chock, 2020; Kirchgaessner, 2021a; 2021b; Lewis & Hilder, 2018; Naughton, 2022). More recently, research distilling the growing range of AI ethical principles suggested that, in general, the principles of beneficence, non-maleficence, autonomy, justice, and explicability are present (Floridi & Cowls, 2022), but that there remains confusion over how best to translate these principles into practice. Despite the number and variety of guidelines published, there is a lack of consensus regarding who is responsible, how, and for what,

☆ This work was conducted with the financial support of the Science Foundation Ireland grant 13/RC/2094_P2 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero - the Science Foundation Ireland Research Centre for Software (www.lero.ie). This work was also funded by the UKRI EPSRC grants EP/R013144/1 and EP/T017465/1. The authors also wish to thank Praveen Kumar Telugu for helpful discussions.

* Corresponding author.
E-mail address: Sarah.robinson@ucc.ie (S. Robinson).

https://doi.org/10.1016/j.jrt.2024.100087

Available online 4 June 2024

when considering software-related harms (Munn, 2023).

Even across documents where consensus appears in terms of principles there are conflicting definitions and definitional tensions. For example, the Montreal Declaration of Responsible AI [(Université de Montréal, 2018), p. 8] understands beneficence as 'well-being of all sentient creatures,' which Munn suggests could be interpreted by technologists as enabling greater connection and communication amongst humans, despite the huge carbon and environmental cost (Munn, 2023). Similarly, the European Commission's Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems, states that autonomous systems, 'must not impair [the] freedom of human beings to set their own standards and norms.' [(European Commission, Directorate-General for Research & Innovation, 2018), p. 16] And in general, there is a view that humans should hold onto the power to decide which decisions to take and which decisions to delegate to autonomous machines. Which humans have this power, however, is less clear, and remain a real concern. Concepts of justice inherent in guidelines often relate to the idea of fairness and distributional justice. Whilst fairness and distributional justice are important, we feel they do not sufficiently tackle the injustices evident in processes of oppression such as marginalization, exploitation and cultural imperialism mediated by software enabled systems. Existing guidelines also fail to consider problems with existing social norms, nor do they question taken-for-granted consensus where it does exist.

Recent work has tried to move from principles to practice, and from siloed actions (often technical in nature), to consider a socio-technical systems wide approach. For example, Nabavi and Browne (2023) sensitize us to the importance of leverage zones within a socio-technical system where practical change can be enacted. These zones range from tackling the system's purpose – the norms, values and goals of developers that get embedded-, to the parameter zone which focuses on technological solutions to design-related harms. Others adopt an eco-system perspective, either through examining the expectations and technological frames embedded in the EU's approach to build a responsible AI eco-system (Minkkinen et al., 2023), or in terms of considering meta-responsibility, the idea that AI systems could become responsible, if they met criteria that enabled answerability across the system (Stahl, 2023). In this paper, we seek to contribute and engage with these more ecological approaches to responsibility, by adding an understanding of responsibility grounded in social connection and infrastructure, which we call *infrastructural responsibility*. To do so, we draw on Star and Ruhleder's proposition that infrastructure is:

> both relational and ecological – it means different things to different groups and it is part of the balance of action, tools, and the built environment [(Star, 2015a), p. 377]

Star goes on to suggest that infrastructure is in fact built into the very fabric of technical work, and if we:

> study a city and neglect its sewers and power supplies (as many have), and you miss essential aspects of distributional justice and planning power….Perhaps if we stopped thinking of computers as information highways and began to think of them more modestly as symbolic sewers, this realm would open up a bit. [(Star, 2015a), p. 379]

Through an infrastructural exploration of responsibility, we seek to open up the Responsible Software Engineering (RSE) space to move away from a focus on responsibility as liability, which isolates responsibility to practitioners and organizations who have acted contrary to norms or principles. In its place, we seek to spread responsibility to everyone engaged in the software development cycle, including users and non-users impacted by infrastructural practice. Through bringing Star and colleagues seminal work from STS on infrastructure (Star, 1990), together with Young's social connection model of responsibility (Young, 2006), we create a new concept relevant for software engineering (and other professions and disciplines) called *infrastructural justice* which we define in Section 2. Through doing so, we make explicit the ways in which our contemporary norms can (sometimes unwittingly) result in irresponsibility and define key leverage points for responsibility that are infrastructurally based and mediated by power, privilege, interest and collective ability. Secondly, we seek to highlight the changing face of software engineering practice through open-source software landscapes of practice. Through doing so, we seek to position open-source software communities as a vital site of exploration and enactment of RSE.

We start, in Section 2, by reflecting on an awareness-raising paper from within SE, which identifies the need for a move toward Responsible Software Engineering (RSE) to extend software quality toward societal impact, trustworthiness, transparency and fairness (Schieferdecker). We then propose a novel analytical lens on infrastructural injustice for RSE research and practice that combines Young's Social Connection Model of Responsibility (Young, 2011) and Star's ecological approach to Information Infrastructures that connected technology and experience (Star & Ruhleder, 2016). In Section 3, we use the conjunction of Young and Star to offer an understanding of responsibility rooted in social connection that is infrastructurally mediated. We apply this theoretical lens to the case study of Deepfake Technology in Section 4, and finish with a call to action for software engineering more broadly to adopt and apply this infrastructural lens through enacting responsibility at key infrastructural leverage points.

1.1. Highlighting responsibility in software engineering

In response to software permeating so many aspects of society that its trustworthiness is more a matter of acceptance than technical quality, Schieferdecker (2020), puts forward RSE as a new frame for SE that attunes to this new context where the role of software has become a matter of public interest. She argues that software transparency, traceability, and explainability have become necessary for users to understand the potential impacts of specific software systems. She also argues that RSE should be constituted by a set of value concerns such as sustainability-by-design, techno-social responsibility and responsible technology development that is based on state-of-the-art SE, and a professional-ethics oath. By techno-social responsibility Schieferdecker indicates that the software community should understand how business models' impact society and in response, they should develop models and infrastructure according to agreed societal principles. Schieferdecker's paper also promotes the importance of professionalisation and stresses the relevance of corporate contexts in the promotion of values such as privacy, security, safety, quality, societal impact, transparency, fairness, trustworthiness, and sustainability.

Unsurprisingly, software transparency, traceability, and explainability are also included as key requirements in the EU Ethics guidelines for trustworthy AI ([HLEGAI] High Level Expert Group on Artificial Intelligence, 2019), and in many of the other AI guidelines (Floridi & Cowls, 2022). Technically and socially, these types of non-functional requirements are challenging to satisfy. Taking explainability as an example, Chazette, Brunotte, and Speith (2022) develop a detailed analysis of how explainability, - which they see as a means of achieving transparency, accountability and trustworthiness – can, with the aid of a set of artefacts they develop, be refined from high level concepts to concrete design choices, with associated methods and metrics for evaluation.

Schieferdeckers' paper is an important contribution in bringing responsibility to the attention of software engineers and associated professional organisations, and the work of Chazelle and others on associated Requirements Analysis testifies to the attention being given to achieving some of the objectives that Schieferdecker identifies for RSE. However, both projects are a step away from people using the technology, still lean on ethical and related principles as a substitute for practice, and do not address cultural practices and norms that might unwittingly lead to irresponsibility, injustice, and harm. Schieferdecker

mentions infrastructure and understands it as a thing, that is strategically created and forged, rather than Star's more ecological and relational understanding, that brings to the fore the importance of social arrangements. Schieferdecker also focuses largely on the corporate context, missing the increasing relevance of engineering practices that take place in open-source communities. Given the ubiquity of software and software-based technology, any approach to RSE that is not sensitive to cultural norms, the rich variety of lived experience, and how that experience is infrastructurally mediated is incomplete. We seek to extend Schieferdecker's concept of RSE, to include a reflective focus on the norms within software engineering practice that create injustice and also offer our understanding of some of the consensus background conditions that need to be understood and engaged with to tackle harms.

1.1.1. Software engineering norms relevant for RSE

Whilst dominant norms such as the move-fast and break things culture of Silicon Valley have been widely critiqued (de Saint-Laurent & Glăveanu, 2023), we want to focus on norms specifically linked to responsibility within software engineering practice.

Norms associated with responsibility in computing and SE are often grounded in concerns about liability (Gotterbarn, 2001; Nissenbaum, 1996). This liability model isolates responsibility to individuals or organizations considered blameworthy for some action or inaction, and can lead to defensive software practice (Gotterbarn, 2001; Nissenbaum, 1996; Thimbleby, 2021; Young, 2011), and a failure to learn from technical or socio-technical errors. For example, deflection of responsibility may be attributed to the 'many hands' problem (Nissenbaum, 1996), where there are so many actors involved in design, development, and implementation, that it is not possible to attribute responsibility to any one person or group. SE is also typically understood normatively and legally as an 'ethically neutral activity' (Gotterbarn, 2001). This norm can lead software engineers to deny professional responsibility when things go wrong, leading, in Gotterbarn's view, to 'a failure to realize that computing is a service to the user of the computing artifact' [(Gotterbarn, 1995), p. 224]. For example, there is a trend of 'written license agreements that accompany almost all mass-produced consumer software which usually includes one section detailing the producers' rights, and another negating accountability' [(Nissenbaum, 1996), p. 11]. As a result, users can be scapegoated when things go wrong (Elish, 2019), (when the fault might be poor design or engineering) (Thimbleby, 2021). This liability culture also means that there is a missed opportunity for reflexive practice to interrogate and learn from errors, as has been undertaken successfully in other industries such as aircraft design and operation, and clinical medicine. Clearly, the organizational context in which software is designed, developed, and deployed is important in the framing and enacting of responsibility, and in creating the norms and practices that contribute to it.

Within SE, there is also very often an emphasis on deferring responsibility to line managers for specific tasks (Gotterbarn, 1995). This leads trainee designers (Gray et al., 2021) or engineers (from our own teaching experience), for example, to make statements like 'I have to do what I get paid to do' when asked or tasked with unsavoury assignments. Task based responsibility can also lead to a narrow problem-solution focus that blinds the engineer to contexts of deployment. Gotterbarn (2001) gives the historic example of the engineer tasked with developing a programme to move an x-ray device above a table to various different positions. The engineer successfully writes the programme to raise and lower the device. However, a patient was later crushed, when a technician set the device to 'table-top-height,' illustrating that the requirements gathering failed to fully take into account the user and the context. In this instance, the machine performed as designed, but the design failed. A narrow emphasis on problem-solution meant that the program worked at moving the x-ray device, but it did not have adequate safety features as they were not prescribed in the original problem.

Whilst ethical neutrality, task responsibility and understanding responsibility as liability all distance the engineer from the user, another important practice that Peterson et al. (2023) suggest leads engineers to psychologically distance from people impacted by software is "abstraction" created by technological mediation and computational thinking. Peterson et al. try to counter this through reflexive case studies aimed at reducing the psychological distance in their teaching, but to our knowledge, this type of approach to educating SEs is not the current norm.

In this section we have introduced Schieferdecker's approach to RSE, which offers a good basis for the development of RSE grounded in principles and professionalism. However, Schieferdeckers' is a generic account with little reference to specific systems, organisations, or cultures. As the evidence above demonstrates, cultural and organisational factors play a significant part in the organisation of SE and in the ways in which technology is used in organisations and institutions, suggesting the need for an ecological approach that takes into account software-in-context as a socio-technical concern. In the next section, we introduce Star and colleagues' understanding of infrastructure which situates professional norms and practices in a wider ecological context; one that is culturally and organisationally sensitive for our approach to RSE.

2. Infrastructure

When we consider many of the harmful impacts of software, such as, the death of teenager Molly Russell, partially attributed by the Coroner, 'to the negative effects of social media', (Leveson, 1995; Walker, 2022) or the wrongful prosecution of more than nine hundred innocent employees in the British Post Office scandal/ IT Horizon scandal for theft and fraud, due to accounting problems with Horizon software (Minkkinen et al., 2023) we need to consider the socio-technical ecologies of injustice to understand where responsibility lies, and how injustice is created. To do so, Star and colleagues remind us of the importance of ecological approaches, and 'refusing social/natural or social/-technological dichotomies.' [(Star, 2015b), p. 15]. In this way, Star's ecological approach is synergistic with the growing body of work that explores the role of eco-systems in understanding responsibility for AI (Minkkinen et al., 2023; Nabavi & Browne, 2023; Smolka & Böschen, 2023; Stahl, 2023; Stahl, 2022). What Star and colleagues (Star & Bowker, 2006; Star & Ruhleder, 1994) add is that ecologies contain infrastructure, which is relational, rather than something universally experienced or a thing like a railway track. In doing so, they ask *when* something is an infrastructure, rather than *what* is infrastructure, and bring into focus, the entanglement of material and social relations that matter for responsibility. Infrastructure cannot be understood without consideration of the people who design, maintain, and use it. In short, it cannot be 'stripped of use' [(Star & Ruhleder, 1994), p. 113]. When we consider software systems as infrastructure, relationality can become particularly visible in the barriers presented to certain groups such as the poor, the disabled or those without access to hardware or training (Star & Bowker, 2006). For example, without a voice-based or braille system, the internet does not support a blind person's communication (Star & Ruhleder, 2016). Thus, we are always talking about the positions offered, constrained, and enabled by infrastructure, *who* that infrastructure is for, and who it excludes.

For Star and Ruhleder (2016), infrastructure emerges with certain dimensions. Firstly, just as machine learning algorithms are often layered within social media platforms, infrastructure is always embedded and 'sunk' into, inside of, other structures, social arrangements, and technologies' [(Star & Ruhleder, 2016), p. 380]. It inherits both the strengths and weaknesses of what Star calls the 'installed base,' a pre-existing structure which acts as a foundation, and opens possibilities, yet can create blocks (conceptual or concrete) for the designer/developer. Infrastructure, as such, does not just emerge 'de novo' [(Star, 2015a), p. 382] but rather is both enabled and constrained by 'the organizational, institutional, regulatory, sociotechnical arrangements that are already in place.' [(Aanestad et al., 2017), p. 29] For example,

much software is based on binary Boolean logic, despite the potential for a counterpart using base 3. Finding alternatives to the binary Boolean logic, however, is currently unlikely, due to the 'immense installed based experience available to the binary code.' [(Herbig & Kramer, 1993), p. 44]

The installed base is a conceptually useful tool which brings our attention to the way in which SE 'entails engagement in processes of extension, recombination, substitution of parts and arrangements that already exist.' [(Aanestad et al., 2017), p. 29]. It also suggests that norms and practices are created through interactions with the installed base and so it is an important analytic tool to consider for RSE. Examples of installed bases that create injustices include the widespread use of open-source code which can create cybersecurity vulnerabilities (Klappholz, 2024; Synopsys Inc, 2024), historic examples such as the way in which Therac-6 and Therac-20 (where bugs in that software had gone undetected due to hardware safety mechanisms) acted as an installed base for Therac-25 and led to the overdose of six patients in 1985–1987. These hardware mechanisms were not continued for Therac-25 due to an overconfidence of the safety of the software. Addressing the errors in the installed base and supporting some of the strengths of it (hardware safety mechanisms) would have been a responsible action in the Therac-25 design which was missed (Gotterbarn, 2001). In a vastly different example of an installed base, the LAOIN-400 m database scraped from a 'billion-sized datasets scraped from the Internet' ' [(Birhane et al., 2021), p. 1] acts as an installed base for deepfake technologies, as well as large language processing models which we explore more in Section 4. LAOIN-400 m contains a high proportion of Not Safe For Work (NSFW) images of sexual violence, and discriminatory annotated search engine terms to fetch images. Thus, the database, as installed base has the potential to perpetuate technological mediated violence against women in the technologies that it informs (Birhane et al., 2021). Thus, the database, as installed base has the potential to perpetuate technological mediated violence against women in the technologies that it informs (Birhane et al., 2021). Correcting problems with the database, would lead to better arguably more just, outcomes.

Infrastructure also supports tasks invisibly, only becoming visible upon breakdown. Infrastructure also embodies standards and plugs into other infrastructures and tools in a standardized way, and similarly is shaped by and shapes conventions of practice. For example, as Chat GPT, a large language model-based chat bot developed by OpenAI, grows in use, it is changing education conventions around academic integrity and academic assignments, and in turn, Chat GPT changes, as users help train it further (Williamson et al., 2023). However, importantly, infrastructure is also a balance point between globalized standardization, and the enablement of local practices, and importantly is characterized by a reach and scope that is temporally beyond a single event, and spatially beyond a 'one-site practice.' [(Star & Ruhleder, 2016), p. 380]. Infrastructure is also learned as part of Membership, or as part of a community of practice, whether that is as part of an open source deepfake technology community, as we discuss in section 4, or as part of formal software training.

Star's infrastructural approach to SE makes clear the relational quality of interactions between people and technology in these contexts. It also makes visible how the installed base enables and constrains action. However, the infrastructural approach alone does not offer a sufficient conceptualisation of responsibility. Young's social connection model of responsibility provides an approach to responsibility that complements Star's infrastructure. To advance our understanding of responsibility we need to understand how Young and Star could interact with each other in a manner that brings into focus the concept of infrastructural injustice, which addresses what RSE could be responsible for. Some of this interaction is methodological, and some of it is conceptual. Before, we consider infrastructural injustice, however, we must first understand what we mean by Justice, and for this we once again turn to Young.

2.1. Infrastructural justice, and socio-structural processes

Star argues, as we mention in the Introduction, that an infrastructural focus is important to make visible essential components of distributional justice that could otherwise be overlooked. Often when we consider justice, we are similarly considering distribution, that is the just or fair distribution of goods, services, opportunities and resources in society. Yet Young helps us realise that whilst (re)distribution is central to justice, redistribution alone is insufficient to achieve it. Take economic domination for example. It could be understood as a simple matter of some people in society having more money than others. However, for Young, 'economic domination derives at least as much from the corporate and legal structures and procedures that give some persons the power to make decisions about investment, production, marketing, employment, interest rates, and wages that affect millions of other people. Not all who make these decisions are wealthy or even privileged, but the decision-making structures operates to reproduce distributional inequality and the injustice constraints on people's lives.' (p. 23). In response, Young's concept of justice focuses on the 'institutional conditions that make it possible for all to learn and use satisfying skills in socially recognised settings, to participate in decision-making, and to express their feelings, experiences, and perspective on social life in contexts where others can listen.' (p. 91). In short, for Young, justice is about self-development, and self-determination, and achieving justice is about alleviating oppression and domination.

Young argues that whilst all forms of oppression involve domination, not everyone who is subjected to domination is oppressed. Thus, oppression is much more than unequal access to economic resources, but is for Young inherently related to decision-making power, the social division of labour, and cultural and normative factors that create exploitation, marginalization, powerlessness, cultural imperialism and violence.

These five 'faces' of oppression have different flavours and Young explains that exploitation, marginalisation and powerlessness are processes linked to the unequal division of labour in society, and to how certain groups or people are recognised or not. For example, exploitation involves social processes where the outcome of the labour of some, creates profits for others and often intersects with race and gender. A contemporary example of exploitation is evident in the data annotation practices of companies like Open AI who have historically outsourced data labelling (sometimes of illegal and explicit content) to companies with employees based in Kenya, India and the Philippines (Rowe, 2023). Data annotators in these contexts report being paid poorly and being exposed to damaging online content of sexual violence and abuse without adequate psychosocial support. Marginalisation refers to the processes whereby people are 'othered' and deemed useless to the system or infrastructure. These social processes often impact older adults in the Global North, disabled people and people who are not considered 'productive'. We see marginalisation in the gross language disparities in natural language processing technologies, where 2191 languages spoken by 1 billion people, representing almost 90 % of global languages are not represented in any NLP technology, potentially because there is no 'market' for technologies in these languages (Joshi et al., 2021). This contributes to a growing digital divide, and contributes to hegemonic language practices that obscure important local languages. Powerlessness involves exploitation hierarchies, where capitalist owners exploit professional classes, and in turn those professional classes gain privilege and benefits from the ongoing exploitation of the working classes. Powerlessness in this way is linked to the ability to define the task or job which is central to justice for Young, and an issue we see in the outsourcing of software development work such as annotation we discuss above, but also coding itself by clients who set the parameters in freelance software arrangements (Rauf et al., 2023).

The final two 'faces' of oppression can also be linked to material distributional aspects of justice, but Young suggests they are more structural and cultural, and once again, we will argue infrastructural.

These are cultural imperialism and violence. Cultural imperialism refers to the 'universalization of a dominant group's experiences and culture, and its establishment as norm' in a way that leads to a process of 'othering' and rendering invisible other groups preferred representations and forms of knowing. (p. 59) This form of oppression Mejias and Couldry suggest is omnipresent in the data grabs or 'the continuous extraction of economic value from human life through data' [(Couldry & Mejias, 2023), p. 787] that they consider as a continuation of colonialist extraction, and which disproportionately impacts minorities and populations in the Global South. We have already mentioned the problems with large databases used to train LLMs (Large Language Models) and the misogynistic representation of minority women within them. Khan and Hanna demonstrate how AI technologies trained on similar databases, are also increasingly involved in injustices, from the misrecognition of minority faces by facial recognition software in the United States, to the mistranslation of captions by Facebook in Palestine (Khan & Hanna, 2022). This brings us to Young's final face of oppression – violence, which is physical and or psychological harm towards a person or group, as well as harassment and intimidation. Such violence is perpetuated more easily through contemporary AI enabled weapons, but less overtly in deepfake pornography which we will detail more in Section 4.

Whilst so far, we have introduced Young's theory of justice, we want to make Young's theory infrastructurally relevant for software engineering. To do that, we combine Star's infrastructure with Young's ideas around justice, to first develop a concept of infrastructural injustice, that is:

> a kind of moral wrong distinct from the wrongful action of an individual agent or the repressive policies of a state (or institution). Infrastructural injustice occurs as a consequence of many individuals and institutions acting to pursue their particular goals and interests, for the most part within the limits of accepted rules and norms [(Young, 2011), p. 54] (*infrastructural added to replace structural, and (or institution) added to state*).

Young (2011) argues that this type of injustice arises from four socio-structural processes. Firstly, the norms and practices that software engineers take-for-granted can contribute to injustice and 'produce and reproduce' the infrastructure the facilitates injustice (McKeown, 2021). For example, by drawing on 'task responsibility', complex contexts of deployment are ignored as we have demonstrated in the Therac-25 example above.

Secondly, norms and practices include 'the unintended consequences of the combination of the actions of many people' [(Young, 2011), p. 53] which can include 'deprivation and oppression for the least advantaged people in the structure.' [(McKeown, 2021), p. 3].

Thirdly, legal and social rules are experienced as social facts and combined with the material world, constrain action the installed base. For example, capitalist business logic as social fact means that even though social media companies are reportedly aware that engagement-based algorithms create division, worsen youth mental health and spread mis- and disinformation (Amnesty International, 2022; Barger et al., 2016; Golbeck, 2020; Naughton, 2022) etc., companies are unwilling to change back to chronological news feeds, for fear of losing competitiveness and profit. This, in turn, impacts the type of SE that social media companies require, and shows how SE norms interact with norms from other industries, disciplines and society to create injustice (s).

Lastly, Young suggests that a fourth socio-structural process relates to infrastructure as macro social space, in which people occupy related positions, which enable and constrain interactions. For example, infrastructures both include and exclude and offer positions for action. They also enable what can and cannot be said. For example, Geoffrey Hinton, a leading cognitive psychologist, and computer scientist, who drove advancements in deep learning and worked at Google Brain, recently resigned, stating, 'If I'm employed by Google, I need to keep thinking, how is this going to impact Google's business?' (Hern, 2023). Thus, his institutional position limited what he could and couldn't say, and by changing position, he is freed up to critique technology advancements.

These socio-infrastructural processes tell us about the conditions necessary for infrastructural injustice to occur, they do not, however, point us to how to address injustice and counter it. Young's social connection model of responsibility will help us to do this.

3. Infrastructural (In)justice & the social connection model of responsibility

Young suggests that everyone who participates in infrastructure is responsible for injustices created within it. She argues that even when specific institutions or individuals are found liable or causally responsible for a harm, this does not absolve others from responsibility to ensure the infrastructure becomes just. At a foundational level, this means being able to take the position of the other. Whilst the spreading of responsibility is often cited as a problem in SE e.g. the many hands problem (Santoni de Sio & Mecacci, 2021), Young supports *recognition of responsibility* that is both individual and shared. This includes those with *power* who have the capacity to influence structural processes, and those with *privilege*, who benefit from the structural exploitation of others, as well as the exploited who often have the most interest in pursuing change. By emphasizing that the exploited are also responsible for software related harms, our reading of Young implies the importance of starting with *felt experience* of what it is like to be exploited infrastructurally. For example, in her work on fast fashion and the global sweatshop movement, Young cites how non-government organization (NGO) campaigns aimed at addressing structural injustice failed, as they did not consider the daily realities of sweatshop workers, whose lives were in fact worsened by the NGOs well-intended actions.

For Young, everyone who produces and reproduces the structure through their practices and actions is responsible in a 'partial way' [(Young, 2010), p. 380]. In Myanmar, Facebook has been accused by the Rohingya minority of spreading hate speech and algorithmically supporting genocide. Examining this claim through Young's lens might bring us to examine the partial responsibility of the company, the users, non-users impacted by the platform, the product managers, the engineers designing the algorithms and maintaining them, the content moderators amongst others are all responsible.

Secondly, and as we have argued throughout, Young also turns our attention to the harm created by everyday taken-for-granted norms and practices. In other words, through our actions we often (unwittingly) create harm and injustice, and it is consensus that is the problem that is rarely probed. Within SE as we have described in Section 2, there are norms, conventions, standards, and objects that are naturalized (Dittrich et al., 2020; Sharp et al., 2016) and sometimes harmful, as illustrated by the open-source code vulnerabilities example above. Star reminds us that these are not always predetermined but are naturalized through practice over long periods of time. The more naturalized the object, the more 'it sinks into the community's infrastructure' [(Star, 2015c), pp. 153–154] and becomes invisible.

Young and Star both encourage us to make this naturalization process visible and to study it. This might mean a renewed RSE focus, for example, on software architecture and requirements analysis and the ways in which certain objects such as documentation standards with guidelines for structuring requirements documents become unquestioned (Nuseibeh & Easterbrook, 2000). To realize responsibility, Young indicates we move away from blaming individuals (often as we have demonstrated in section 2 the user), to question these norms, rules, and practices, and be explicit in our reflection and deliberation on their impacts. Young centres the importance of being reflexive, that includes holding 'a mirror up to one's own activities, commitments, and assumptions'…. being 'mindful that a particular framing of an issue may not be universally held.' [(Stilgoe et al., 2013), p. 1571]. Thimbleby

(2021) suggests that there is a marked absence of reflexivity in digital healthcare due to the constraints of liability, and task responsibility. Reflexivity can also be constrained by client demands, but Young suggests that we remain responsible even when faced with these demands.

Thirdly, Young's model also emphasizes the importance of forward-looking responsibility, which addresses the 'ongoingness' of injustice and the need to mitigate future impact. This is different from liability. The model understands injustice as occurring often in unintended ways, and that blame in this sense would not be productive or sufficient. Young is concerned with mobilizing everyone who contributes to the infrastructure 'to transform those processes' [(Young, 2011), p. 109] in future-oriented behavior. However, to do so, she acknowledges the importance of understanding how infrastructural injustice is produced. The aim of reflexivity is to make these linkages visible. Forward-looking responsibility means embracing methods for envisioning direct and indirect societal impacts of software that go beyond the immediate design brief, and consider for example, if the software becomes pervasive and ubiquitous as well as indirect and direct stakeholders (Nathan et al., 2008). In our instance, this would mean examining the installed base for infrastructure, to look back at what elements might contribute to irresponsibility and injustice if reified infrastructurally. However, it also means looking ahead; understanding how requirements and/or architecture embed certain possibilities and constrain others, whilst also thinking about what actions are needed now for a responsible future.

Lastly, starting with 'experience as experienced' (McCarthy & Wright, 2023) is the beginning of collective action. Young locates structures as sites of collective action. Young is less clear on how that collective action should take place or how people should organize, but she stresses that those who are aware that injustice occurs must persuade others of this fact, and through making the background conditions visible, collective action and change can occur.

Young's central thesis is that all those who participate in some way in structural processes bear some responsibility to make them just, but we are not all responsible equally. Young offers four parameters of reasoning (power, privilege, interest, and ability) to effect change and to address infrastructural injustice. Star, however, would suggest that reasoning does not just happen cognitively, but is mediated by the characteristics of infrastructure. Bringing Young and Star together again supports us to address who is responsible and in what way for infrastructural injustice.

3.1. *Infrastructural justice and Young's parameters of power, privilege, interest and collective ability*

3.1.1. *Power and privilege*

According to Young, people with power have a greater responsibility and capacity to try to make infrastructure just, but they also have many privileges to lose. However, when we benefit infrastructurally it is often difficult to see the harm that infrastructure facilitates. The powerful and the privileged 'must usually be pressured to take next steps, aimed at changing the effects of their actions.' [(Kasirzadeh, 2022), p. 9] Other people with privilege do not have institutional power to change the infrastructure. Despite this, their position of privilege results in a greater responsibility in organizing efforts to correct injustice, because they will not suffer 'serious deprivation' from the change that occurs (Young, 2011).

For Star, power is about whose master narrative is dominant (Star, 2015a; Star, 1990), and she explores ways in which standards and standardized practice works to marginalize, exclude or silence. In this way it is not just people who have power, but it is the relationship of people with the materiality of the infrastructure that enables and constrains. For example, Project Maven, a collaboration between Google and the United States Department of Defence to augment the 'vision' of drones using AI, met with the power of staff with niche skills, who refused to develop the necessary 'air gap' security technology. Google were forced not to renew the military contract (Boag et al., 2022) which could have resulted in Google's technology contributing to lethal force. This example illustrates infrastructure is always relational and enabled or constrained by the actions of many working in tandem.

3.1.2. *Interest and collective ability*

There may also be those without power and privilege within an infrastructure, but through their interest in making it just, share in the responsibility to do so. This could be allies, or those directly or indirectly impacted, which Young defines as *interest*. A recent example of interest is the campaign led by the father of British Teenager, Molly Russell following her death, actively pushing for social media platforms (Walker, 2022) to do more to keep children safe online. Interest is often an obvious reason for trying to make infrastructure just, yet it is not often achieved with interest alone, and alliances across the structure with more powerful or less vulnerable allies can be helpful in actioning responsibility.

Following on, a fourth parameter of reasoning is *collective ability,* or the resources or capacities required for collective organization, that might be a trade union, a counter-public or some other infrastructural form that could support agitation for justice. The refusal of core Google staff to provide the skills necessary to complete Project Maven discussed above, was preceded by an open letter of over 4000 google staff protesting the company's decision to collaborate with the American Military. This led Google management (those with power) to reconsider Google's strategic direction and culminated in a set of AI principles which now prevents Google from engaging in lethal force (Boag et al., 2022).

Infrastructure itself also enables and constrains collective ability. For example, Reddit recently acted to prevent free data scrapping for training tools such as GPT-4. Reddit acted to monetize data, suggesting that 'it must be fairly paid to continue supporting high-usage third-party apps.' (Naughton, 2023) As a result, Reddit is now charging for its Application Programming Interface (API) which is used by many apps, but also by many Reddit Users and moderators (including blind users), to access the site in an accessible way on mobile phones. In protest, many sides including r/software have mobilized a black-out (r/software, 2023).

Whilst collective ability should not lead to a privileging of infrastructural issues over others, it remains an important leverage point for enacting responsibility. Sometimes infrastructure can itself create silos where collective ability becomes limited or challenging, for example, the type of infrastructural injustice created by social media's engagement-led algorithms arguably limits the self-development and self-determination of - young people with mental health difficulties (Naughton, 2022), diverse actors impacted by the spread of mis and disinformation (Johnson et al., 2022), minority and marginalised groups impacted by the amplification of hate in places of linguistic diversity where safety mechanisms are not in place, (Amnesty International, 2022) and yet these 'consequences' are siloed (Backhouse et al., 2023). Rarely do we see the various interest groups join forces to champion 'infrastructural' change that would cut across topics. This might be due to a lack of 'infrastructural awareness' as well as what Gran et al. label algorithmic awareness (Gran et al., 2021).

Young also illustrates that if one of the four parameters (power, privilege, interest, and ability) is missing, often those in power can block progress, those with privilege can protect vested interests, and those with experience of exploitation might not have the resources or safety to take a stand. However, for Young, 'the infrastructural processes can be altered only if many actors in diverse social positions work together to intervene in these processes to produce different outcomes' [34:123] (infrastructural added to replace structural). We propose that for RSE to be realized, these four parameters of reasoning must be considered along with the infrastructural ways in which power, privilege, interest, and ability can manifest.

4. Application of an infrastructural-responsibility approach to deepfake technology

In this section, we will apply the infrastructural-responsibility approach to a case study of deepfake technology, to illustrate the utility of the model, which must always be understood in its context of application. Deepfake technology uses a form of AI called deep learning to edit video files, e.g., replacing the face of one person in the video with that of a different person (Sample, 2000). In this case, forward-looking responsibility relates to the ongoing injustices created, particularly towards women, through deepfake non-consensual porn (an estimated 96 % of all deepfake instances (Ajder et al., 2019)). Other responsibility concerns relate to the spread of fake news and fake evidence in Court (Westerlund, 2019), and how the combination could further erode public trust (Fallis, 2021). Deepfake technology has potential to impact political and democratic processes (Ajder et al., 2019; Pawelec, 2022), terrorist and rogue state propaganda (Purwadi et al., 2022), market manipulation (Wilson et al., 2020), blackmail (Lucas, 2022) and to help people commit fraud (de Rancourt-Raymond & Smaili, 2023). Many of these uses have already occurred, yet there is a lack of concentrated collective action to address them. Importantly, harms associated with deepfake technology can be considered a form of infrastructural injustice, in that deepfake technology is not always created by malicious actors but can also be created by the norms and everyday taken for granted practices of open-source communities who wish to learn how to "do" AI. It also contributes to violence against women, and a form of cultural imperialism whereby women are sexualised and exploited without their consent.

Whilst our infrastructural justice approach has many components, we have decided to focus on four elements of Young's background conditions and Star's infrastructure for illustrative purposes below. These are the fact that infrastructure is: (1) Learned as part of membership; (2) Links with Conventions of practice; (3) Embodies Standards and (4) is based on an installed base, and the ways these four background infrastructural conditions interact with power, privilege, interest and collective ability.

4.1. Learned as part of membership

Deepfake technology is opensource software (OSS), the development of which was supported by social media platforms such as Reddit, repositories on GitHub, and data repositories/datasets such as LAION-400 m (Ajder et al., 2019; Gamage et al., 2022a; Johnson & Diakopoulos, 2021; McCosker, 2022). Being part of this community of practice means learning deepfake techniques from more experienced coders, who (generally) collegially impart skills and knowledge (Newton & Stanfill, 2020). Learning resources include tutorials by AI influencers, tips, and techniques on sites such as GitHub and/or Kaggle.com, and via YouTube tutorials, Discord channels and subreddits etc. Research suggests, however, that the community is dominated by a norm of 'toxic geek masculinity,' created by men traditionally marginalized within mainstream hegemonic masculinity and who exert dominance upon women technologically, with no regard for the consequences, or for consent (Newton & Stanfill, 2020). This culture is maintained through norms and practices that enable psychological distancing strategies i.e., the most common term used to refer to human subjects is 'face' or 'data' and there is a distinct absence of discussion of impact (Newton & Stanfill, 2020).

In addition, there is a neutral stance towards pornographic content, which is prized for 'the technical knowledge communicated alongside it' [(Newton & Stanfill, 2020), p. 404] rather than acknowledged as harmful, and as contributing to the domination and oppression of women. Anonymity strategies are also frequently deployed including redacting computer usernames that would ordinally show when posting a file path or using throwaway accounts. For instance, there are some GitHub accounts that only discuss deepfakes rather than generating code. Qualitative research indicates that most community leaders are concerned about the reputational damage of their software being used for illicit non-consensual purposes (Widder et al., 2022). McCosker (McCosker, 2022) also indicates that many novices are interested in learning how to 'do' AI rather than deepfakes per se, so that harm could be unintended, a core aspect of Young's idea of infrastructural injustice.

The community of practice is not primarily commercial, but some (particularly influencers) do profit from sharing information and have monetized resources, indicating differing levels of power and privilege. Influencers, for example, who train and teach these skills also gain privilege and status (McCosker, 2022). Responsibility washing is also evident: See, for example, the FaceSwap Subreddit ReadMe file that states that FaceSwap should not be used for unethical, illicit, or questionable content, but rather that, 'We will take a zero- tolerance approach to anyone using this software for any unethical purposes and will actively discourage any such uses. They describe the deepfake community as 'a fantastic learning opportunity.' Cited in [(Newton & Stanfill, 2020), p. 405].

4.2. Links with conventions of practice

The 'radical transparency' within open-source code allows the capture of implementation bugs, but paradoxically allows downstream harms of misuse to thrive (Widder et al., 2022), which disproportionately impacts women, particularly celebrity women whose data is widely available, and women who work in porn (Ajder et al., 2019): *Open-source licences* prohibit developers legally and normatively from controlling downstream uses of the technology. This is despite the named 'zero-tolerance approach' stated on the ReadMe files, which is in tension with the infrastructural ability to maintain this approach. The open-source ethos also embodies the idea that *technology is neutral* (like what we have discussed in Section 2.), diffusing the responsibility of developers. Users also have the right to use the code for *any* purpose, as evident in the ReadMe files which denies any responsibility for use cases. However, this masks the power and privilege of coders/developers and undermines their interest and ability to change it. This is both a convention of practice, a standard and part of the 'inertia of the installed base' that we will discuss further below.

Applying an infrastructural lens opens a leverage point of responsibility, regarding what type of licence a developer decides to use when the software is developed. Widder et al. (2022) indicate that leaders of deepfake projects have some discretion here but once a licence is a General Public Licence 'which guarantee[s] end users the four freedoms to run, study, share and modify the software' (Wikipedia, 2023), there is no going back. Even if it were legal to do so, given the distributed nature of open-code software, it would be impossible to trace all authors to request their permission to do so. However, developers have agency in the initial licence decision and given that open-source licences can lead to harm downstream, choosing a licence very carefully is an important leverage point of forward-looking responsibility.

Whilst community of practice leaders have said they cannot be 'responsible' due to the legal and normative standards required by Open Source's values and legal obligations, the societal harms created by deepfake technology warrant that the infrastructure is also shaped by *conventions of counter-practice* (Widder et al., 2022), where some members of the community actively make a choice regarding to which projects their labour contributes. There is also a new emerging movement called the Ethical Source Movement which indicates that the Open Sources norm of 'Freedom Zero' which 'is the freedom to run the program as you wish, for any purpose,' is outdated and is creating societal harms (Vaughan-Nichols, 2023). They have developed a form of licence called a Hippocratic licence which states that:

> The Software shall not be used by any person or entity for any systems, activities, or other uses that violate any Human Rights Laws (Vaughan-Nichols, 2023).

They further counter the norm in open-source communities that

technology is neutral stating that:

> We have to accept that our work has an outsized impact on society, and that means we have an outsized responsibility to minimize the harm it can cause (Vaughan-Nichols, 2023).

4.3. Embodiment of standards

The Deepfake infrastructure also embodies the standards of its landscape of practice. The lack of moderation standards on GitHub can lead to implementation harms and the failure of moderation standards on Reddit, mean that some content still signposts novice coders to pornographic content (Winter & Salter, 2020). Whilst Reddit banned all non-consensual porn, in response much of the deepfake community migrated to other platforms and found work arounds for sharing banned content (Winter & Salter, 2020). This illustrates how a responsibility intervention designed to curb illicit use has been ineffective. Reddit now acts as an archive, as members of subreddits have previous information backed up on their computers and when this information is requested share it via email or other platforms (Gamage et al., 2022b). In addition, there is no moderation on GitHub, yet it is the 'space where discourse of Reddit becomes action [(Winter & Salter, 2020), p. 395] and a primary space where deepfake pornography is shared and consumed. We can see how the infrastructure is relational, enabling certain harmful actions. Further still, we can see how it leads to domination and oppression infrastructurally in that women impacted do not have a say in how images about them are used.

4.4. Installed-base

The *installed base* comprises big data often scraped from the internet without the consent of data subjects. As already mentioned, Birhane and colleagues suggests that data repositories such as LAION-400 m have inherited the misogynistic ethnocentric stereotypes from the internet (Birhane et al., 2021). It also comprises deep learning models and Machine Learning combined with social media platforms (with engagement-based algorithms) that enable rapid sharing of deepfake knowledge and distribution of content. Due to advancements in AI synthetic media and GANs, by 2017, Deepfake (initially a code name for a coder on GitHub and Reddit) developed the first pornographic deepfake and started the subreddit r/deepfake, which enabled a community of practice to form. Deepfake infrastructure has inherited the strengths and weakness of this installed base, and the combined entanglement of these various technology developments has enabled the societal harms associated with deepfake technologies to proliferate, indicating the importance of addressing injustices associated with installed bases before they are appropriated for other means.

In Table 1 we illustrate how these elements of infrastructure interact with power, privilege, interest, and collective ability to currently constrain action, and argue that applying this approach can also make

Table 1
Power, privilege, interest & ability: deepfake community of practice.

	Power	Privilege	Interest	Collective ability
Learned part of membership	Developers who benefit commercially AI Influencers Platform owners e.g. Microsoft Social media platforms & their shareholders	Developers who benefit commercially AI Influencers Consumers of Deepfakes Social media company shareholders	Subreddit leaders GitHub collaborators Journalists (due to the undermining of public trust in journalism) Victims of deepfake cloning (often but not limited to porn actresses & celebrity women) Data subjects The general public & public & State actors (due to the impact of erosion of public trust)	Platforms have abilities to bring in infrastructural changes Men who build deep-fake porn could find ways to assert positive masculinity rather than toxic masculinities Deepfake community leaders could challenge norms Novices in CoP who just want to learn how to "do" AI could be supported to act Celebrities have a platform to make this a global public issue and assert pressure infrastructurally
Conventions of Practice	Power is wielded through: Open-Source Software Licences & Norms Norms of toxic geek masculinity	Privilege is granted to those learning "AI" for free who gain from the system	Women (and minorities) as data subjects have an interest in remedying injustice. Gender based violence campaigners have an interest Men as allies have an interest in tackling toxic forms of masculinity. ACM/IEEE communities have interest in understanding how conventions of practice violate their principles People who just want to learn AI	The growing move to have licensing clauses and find counter-practice conventions offers collective ability to remedy injustice Addressing toxic geek masculinity in online spaces offers a challenge for collective ability, and the need for collective action across groups that may not usually collaborate such as gender specialists, and Platforms such as Github
Standards	GitHub, Pornhub and Reddit (among others) do not have sufficient moderation standards Anonymity as a norm/ standard on GitHub	Privileges a lack of accountability for action amongst largely male group of coders	Same as above	Changing standards to ensure traceability
Installed Base	Data repositories e.g., LAOIN-400 m Engagement based algorithms – spread DFs rapidly Deep-learning and Machine Learning techniques The Cloud Platforms Lack of consent laws Deep learning techniques Toxic masculinity as norm Gender based violence Pornography	Data repository managers, staff, shareholders and users of them all have privilege within DF infrastructure Social media shareholders, staff, management also gain privilege from the rapid proliferation of DF on their platforms Cloud companies, staff, shareholders, users benefit from the infrastructure that enable DFs	Same as the above Academics in Data Justice, HCI, Software engineering, information systems & gender studies have an interest in bringing to the fore the problems with the installed base	There is a collective action opportunity to make visible the installed base & to question harmful reifications e.g. of LAOIN-400 m database

visible alliances across power, privilege, interest, and ability that might support responsibility in this illustrative example.

5. Discussion

In this paper, drawing on Young (2011) and Star (2015a), we suggest that responsibility is not blame, but is a reflexive individual and shared process, that is infrastructurally mediated to address instances of infrastructural injustice, which can be achieved by addressing domination and oppression. We emphasized the importance of considering everyday norms and practices within SE as critically important to enacting RSE and tackling the challenges it must address, including infrastructural justice. A focus on norms and practice is not new in SE (Sharp et al., 2016) but has yet to be explicitly prioritized in the RSE agenda. In our view, norms and practices are critically situated within infrastructural processes, making our new concept of infrastructural justice an important consideration for RSE, which offers key leverage points to enact responsibility. This redefines the scope of the problem that RSE must address, and broadens the parameters of social action beyond engineering, design, and development to consider the obligations of people who are infrastructurally connected. This in turn redefines who is responsible in a modern IT society, but the level of responsibility is carefully mediated by *power, privilege, interest,* and *ability*. Collective action is also identified as the primary mechanism through which infrastructural justice is achieved.

Through problematizing norms and practices, we seek to make visible the taken-for-granted ways in which responsibility is constrained, and how through addressing these *background conditions* RSE could be enacted. Whilst infrastructure is always particular, the ways in which Star and colleagues describe its characteristics are universally applicable as an analytic lens to the identification of leverage points for responsibility. Nabavi and Browne have already identified the importance of leverage zones in socio-technical systems for responsible AI and offer five Ps (problem, parameter, pathway, process, and purpose) to sensitize software developers to enacting responsibility (Nabavi & Browne, 2023). We suggest that the purpose of an ecology is often embedded in mundane aspects of infrastructure from standards and conventions of practice to the way in which the infrastructure is learned as part of membership of a community. Instead of asking sensitizing questions of what purpose is this system for, we suggest asking questions about *when* does this infrastructure become visible? *Who* does it include and exclude? What norms does the infrastructure enable? What reach and scope does it have? And what impact does this have on society, on the environment? In doing so, we believe that leverage points of responsibility across an infrastructure will become visible for action.

We also seek to sensitize engineers/developers to the importance of considering Young's four parameters of reasoning with each infrastructural relationship. For example, when we consider the installed base(s) of deepfake technology, we need to ask who is given power and privilege? Who is exploited and marginalized and therefore interested in rectifying injustices created by deepfake infrastructure? And what collective ability is there to act across theses spheres to bring about justice? In Table 1 we have highlighted some of the potential collective action alliances (ability) to support responsibility, these include collaborations across disciplinary silos with potential for gender experts to work together with platforms to address 'toxic geek masculinity,'; Academics and legal scholars interested in AI licensing clauses might come together with open-source communities to address problems with contemporary open-source licenses that lead to downstream harms. We might also see Porn Star Actresses, Celebrity Women and Politicians impacted by deepfake technology making an alliance as data subjects to advocate for data sovereignty, copyright, and consent to be infrastructurally prioritized. This indicates that who addresses responsibility is wider than the RSE ideas that Schieferdecker initially considers and opens collective action possibilities.

Centering collective action and spreading responsibility also has other implications for RSE, beyond the values that RSE proponents have thus far explored. It might, for example, involve encouraging or even designing infrastructures which can support the emergence and functioning of participative publics, that is a 'configuration of individuals bound by a common cause in confronting a shared issue.' [(Karasti, 2014), p. 144] so that collective action can be fostered. Whilst a discussion of designing for publics is beyond the scope of this paper, engineering responses might include requirements eliciting for how to support the formation of publics, that could enable reflexivity on shared interests related to responsibility. A focus on participative publics also brings into focus recursive publics, that is 'a public constituted by a shared concern for maintaining the means of association through which they come together as a public.' [(Kelty, 2013), p. 1] In our paper, we have demonstrated how 'toxic geek masculinity' is omnipresent in the open-source sites that enable deepfake technology. Whilst open-source communities as recursive publics have redistributed power and harnessed much societal good in terms of access to free software, our deepfake example demonstrates that this is not always the case. An infrastructural responsibility approach renders these relatively new organizational and social arrangements as central to understanding how responsibility must be considered in context and poses questions for future research to answer, such as how can a recursive public become responsible? Whilst RSE is traditionally the remit of computer scientists, data analysists, software engineers and developers, the rapid de-professionalization of much software development through open-source communities brings to our attention the importance of new socio-technical arrangements that matter for responsibility and signals the importance of a research agenda that explores the entanglements of traditional software engineering with open-source platforms and communities.

These new socio-technical arrangements also create new landscapes of practice. Vulnerabilities inherited from open-source code, as an installed base for other software products, is now one of the leading causes of cybersecurity vulnerabilities in proprietary code. An infrastructural responsibility approach, at the very minimum, asks that software engineering and other disciplines focuses very specifically on the installed base as a concept for responsibility. Such a focus is already happening for data practices associated with machine learning and large language models, with encouraging results on how injustice is present in contemporary practices, and recommendations to make these systems more just (Bender et al., 2021b; Denton et al., 2021; Hutchinson et al., 2021; Khan & Hanna, 2022; Miceli et al., 2021; Scheuerman et al., 2021). The potential for justice is enormous if problems with the installed base can be tackled in a range of contexts.

To achieve infrastructural justice, there also needs to be a shift from responsibility based on individual normative transgressions, to how consensus-held norms and practices create injustice. Within software engineering cultural norms include, but are not limited to, task responsibility, the likelihood of capital business logic driving the agenda, and a move-fast and break things culture which 'devalues the incremental and cooperative care required to create high quality dataset's' (in the instance of our installed base example above), and 'has been implicated in the discriminatory effects of technological systems.' [(Hutchinson et al., 2021), p. 563]. Harmful norms might also include skipping requirements analysis and jumping straight to implementation or the lack of requirements documents that are currently available for data used in Machine Learning (Hutchinson et al., 2021). Outside of professional software engineering circles, the norms of 'toxic geek masculinity' and of open-source licencing both lead to different types of injustice that disproportionately impact women. A focus on cultures and sub-cultures could open up spaces for RSE that have previously not been considered.

Our starting assumption in this paper is that there is a strong commitment among Software Engineers, to ensure that the software systems they work on at the very least do no harm and, better, do good. Whilst these ideas are evident in the ACM software code of ethics, and

other AI guidelines, their application can be supported by understanding the infrastructural ways in which practices occur. Our focus on justice, and understanding it as the absence of domination and oppression (exploitation, marginalisation, powerlessness, cultural imperialism and violence) brings new demands for software engineers to consider in their practice. This would mean a radical re-shift of the way in which SE is taught, drawing on Peterson et al. (2023) approach to reducing the psychological distance between the engineer and the user, but also to include urgent education initiatives on social practices such as whistleblowing (Hunt & Ferrario, 2022) and on design refusal (Zong & Matias, 2024). However this reach now needs to move beyond the classroom and traditional software engineering training spheres, to the influencers and platforms that house and nurture deepfakes. Software engineering practices is also rapidly changing with freelance software developers now making up sizeable numbers of professionals engaged in practice (Rauf et al., 2023). Understanding how these new social and organizational arrangements impact understandings and realisations of RSE in context is important to achieve justice.

We have also demonstrated, how responsibility must be enacted through collective action. Yet one of the main barriers is that infrastructure only becomes visible upon breakdown, and arguably when infrastructural injustice has already occurred. For example, the datasets that we describe as an installed base for deepfake and for many AI applications 'do not appear unless the AI models which are built upon them break down; that is, they make egregious technical errors or mistakes in classification. Others break down when people who are not part of the majority – including racial, religious, ethnic, caste, and gender minorities, disabled people, refugees and migrants, and others in the Global South – interact with them at points of disjuncture, some of them violent and life-alternating.' (p. 176). If responsibility can only be achieved by making connections visible, then design and research must focus efforts on making tangible the connections we share with those impacted by these infrastructural breakdowns. This might mean adopting approaches like Veinot et al.'s (2018) approach to preventing intervention generated inequalities in health informatics which advocates addressing access, uptake, adherence, and effectiveness inequalities directly before systems are designed. Preventing infrastructural injustice, however, will involve preventing processes that create domination and oppression across software enabled systems. Given that infrastructure is not universal, and depends on use, this approach will also need to be tailored to context. Future research could explore what this looks like in diverse infrastructural contexts of use from the impact of digital health infrastructures on mortality to the impact of engagement based algorithms on adolescent health and misinformation.

An infrastructural responsibility approach also means enhancing critical consciousness amongst the public about the ways in which infrastructure and SE culture can impact social life. Without infrastructural awareness, publics that form around infrastructural root causes of issues will not form. This might mean that Global Citizenship Education, for example, champions links between local and global software impacts., and that Software Engineering as a discipline engages with other Education sectors on creating this awareness (Backhouse et al., 2023).

6. Conclusion

This paper introduces a new concept for RSE called infrastructural justice, grounded in an infrastructural understanding of responsibility that builds on Young's social connection model and Stars relational infrastructure. This approach moves away from the traditional liability model of responsibility and makes explicit the dominant norms within SE that perpetuate a culture that defends, deflects, and diffuses responsibility. Instead, our infrastructural responsibility seeks to spread shared responsibility to everyone within an infrastructure, in a forward-looking way that tackles the ongoingness of injustice through collective action. By making explicit the often-implicit background conditions (norms, practices, and other critical elements of infrastructure such as the installed base, learned as part of a community of practices, standards, and conventions), new avenues and questions for Responsible SE are generated which we hope might support a research agenda that moves beyond the central focus on AI, to be more holistic of software-in-contexts. We also hope that it stems debate on infrastructural awareness amongst software engineers, but also the general public including global citizenship educators who might benefit from collaborating with software engineers on issues of local-global import.

The main contribution of the paper is to offer a fresh lens to the ongoing research and debate about eco-systems and responsibility in AI. We offer an ecological and relational approach, grounded in use and practice that is mediated infrastructurally. Whilst others have offered zones of leverage across a problem-response paradigms, we seek to offer leverage points across an ecology based on infrastructural components of use. We suggest that intervening in important ways on the installed base can offer much leverage for RSE. An infrastructural lens debunks the magical thinking evident in the term AI, and instead breaks it down into its constituent components and emphasises the way in which standards, conventions, installed base, learned as part of a community offers points of intervention. We also highlight the importance of context and the entanglement of open-source with professional software communities in a way that matters for the realisation of infrastructural justice. Finally, we suggest new orientations for software engineers within their current disciplines, to embrace design refusal if software is unsafe, unjust and harmful.

CRediT authorship contribution statement

Sarah Robinson: Conceptualization, Writing – original draft, Writing – review & editing. **Jim Buckley:** Supervision, Writing – original draft, Writing – review & editing. **Luigina Ciolfi:** Supervision. **Conor Linehan:** Supervision, Writing – original draft, Writing – review & editing. **Clare McInerney:** Supervision, Writing – review & editing. **Bashar Nuseibeh:** Supervision. **John Twomey:** Conceptualization, Writing – review & editing. **Irum Rauf:** Writing – review & editing. **John McCarthy:** Conceptualization, Supervision, Writing – original draft.

Declaration of competing interest

None of the authors have any conflict of interest related to this work. Neither Science Foundation Ireland nor the UKRI Engineering and Physical Sciences Research Council had any influence on the development and write up of this work.

References

Aanestad, M., Grisot, M., Hanseth, O., & Vassilakopoulou, P. (2017). Information infrastructures and the challenge of the installed base. *Information infrastructures within European health care: Working with the installed base* (pp. 25–33).

Ajder, H., Giorgio, P., Cavalli, F., & Cullen, L. (2019). *The state of deepfakes: Landscape, threat and impact*. Amsterdam: DeepTracesLab. Sep.[Online]. Available https://regmedia.co.uk/2019/10/08/deepfake_report.pdf.

AlgorithmWatch, 'AI ethics guidelines global inventory. AlgorithmWatch.' 2023. Accessed: May 05, 2023. [Online]. Available: https://algorithmwatch.org/en/ai-ethics-guidelines-global-inventory/.

Amnesty International. (2022). *Myanmar: The social atrocity: Meta and the right to remedy for the rohingya*. London: Amnesty International [Online]Available https://www.amnesty.org/en/documents/ASA16/5933/2022/en/.

Backhouse, C., Robinson, S., & Barton, C. (2023). Making the invisible visible: how forum theatre can reveal the impact of social media algorithms on local and global justice issues. *Policy and Practice: A Development Education Review, 37*. Autumn.

Barger, V., Peltier, J. W., & Schultz, D. E. (2016). Social media and consumer engagement: A review and research agenda. *Journal of Research in Interactive Marketing, 10*(4), 268–287.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021b). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada: ACM* (pp. 610–623). https://doi.org/10.1145/3442188.3445922. Mar.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021a). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623).
Birhane, A. (2022). *The unseen black faces of ai algorithms*. UK London: Nature Publishing Group.
Birhane, A., Prabhu, V. U., & Kahembwe, E. (2021). Multimodal datasets: Misogyny, pornography, and malignant stereotypes. *arXiv preprint*. arXiv:2110.01963.
Boag, W., Suresh, H., Lepe, B., & D'Ignazio, C. (2022). Tech worker organizing for power and accountability. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 452–463).
Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (pp. 77–91).
Charitsis, V., & Lehtiniemi, T. (2023). Data ableism: Ability expectations and marginalization in automated societies. *Television & New Media, 24*(1), 3–18.
Chazette, L., Brunotte, W., & Speith, T. (2022). Explainable software systems: From requirements analysis to system evaluation. *Requirements Engineering, 27*(4), 457–487.
Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press.
Couldry, N., & Mejias, U. A. (2023). The decolonial turn in data and technology research: What is at stake and where is it heading? *Information, Communication & Society, 26* (4), 786–802. https://doi.org/10.1080/1369118X.2021.1986102. Mar.
Denton, E., Hanna, A., Amironesei, R., Smart, A., & Nicole, H. (2021). On the genealogy of machine learning datasets: A critical history of ImageNet. *Big Data & Society, 8*(2), Article 205395172110359. https://doi.org/10.1177/20539517211035955. Jul.
de Rancourt-Raymond, A., & Smaili, N. (2023). The unethical use of deepfakes. *Journal of Financial Crime, 30*(4), 1066–1077.
de Saint-Laurent, C., & Glăveanu, V. (2023). AI makes Silicon Valley's philosophy of "move fast and break things" untenabl. *The Conversation* [Online]. Available https://theconversation.com/ai-makes-silicon-valleys-philosophy-of-move-fast-and-break-things-untenable-218159.
Deshpande, A., & Sharp, H. (2022). Responsible AI systems: Who are the stakeholders?. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 227–236).
Dittrich, Y., Michelsen, C. B., Tell, P., Lous, P., & Ebdrup, A. (2020). Exploring the evolution of software practices. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (pp. 493–504).
Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction (pre-print). *Engaging Science, Technology, and Society (pre-print)*.
Ensmenger, N. (2021). The cloud is a factory. In T. S. Mullaney, B. Peters, M. Hicks, & K. Philip (Eds.), *Your computer is on fire* (pp. 29–50). Cambridge, MA, USA: The MIT Press. https://doi.org/10.7551/mitpress/10993.003.0005.
European Commission, Directorate-General for Research and Innovation. (2018). *Statement on artifical intelligence, robotics and "autonomous" systems*. Brussels: European Commission [Online]Available Https://data.europa.eu/doi/10.2777/531856.
Fallis, D. (2021). The epistemic threat of deepfakes. *Philosophy & Technology, 34*(4), 623–643.
Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. In S. Carta (Ed.), *Machine learning and the city* (1st ed., pp. 535–545). Wiley. https://doi.org/10.1002/9781119815075.ch45.
Gamage, D., Ghasiya, P., Bonagiri, V., Whiting, M. E., & Sasahara, K. (2022b). Are deepfakes concerning? Analyzing conversations of deepfakes on reddit and exploring societal implications. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–19).
Gamage, D., Ghasiya, P., Bonagiri, V., Whiting, M. E., & Sasahara, K. (2022a). Are Deepfakes concerning? Analyzing conversations of Deepfakes on Reddit and exploring societal implications. In *CHI Conference on Human Factors in Computing Systems* (pp. 1–19). New Orleans LA USA: ACM. https://doi.org/10.1145/3491102.3517446.
Golbeck, J. (2020). Optimizing for engagement can be harmful. There are alternatives. *IEEE Intelligent Systems, 35*(4), 117–118.
Gotterbarn, D. (1995). The moral responsibility of software developers: Three levels of professional software engineering. *Journal of Information Ethics, 4*(1), 54.
Gotterbarn, D. (2001). Informatics and professional responsibility. *Science and Engineering Ethics, 7*, 221–230.
D.W. Gotterbarn et al., 'ACM code of ethics and professional conduct', 2018, Accessed: May 01, 2024. [Online]. Available: https://dora.dmu.ac.uk/bitstream/handle/2086/16422/acm-code-of-ethics-and-professional-conduct.pdf?sequence=1.
Gran, A.-B., Booth, P., & Bucher, T. (2021). To be or not to be algorithm aware: A question of a new digital divide? *Information, Communication and Society, 24*(12), 1779–1796. https://doi.org/10.1080/1369118X.2020.1736124
Gray, C. M., Chivukula, S. S., Melkey, K., & Manocha, R. (2021). Understanding "dark" design roles in computing education. In *Proceedings of the 17th ACM conference on international computing education research* (pp. 225–238).
[HLEGAI] High Level Expert Group on Artificial Intelligence 2019 [HLEGAI] High Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy ai*. European Commision. Apr. [Online]. Available https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
Herbig, P. A., & Kramer, H. (1993). Innovation inertia: The power of the installed base. *Journal of Business & Industrial Marketing, 8*(3), 44–57.
Hern, A. (2023). We've discovered the secret of immortality. The bad news is it's not for us": Why the godfather of AI fears for humanity. *The Guardian, London*. May 05Accessed: May 08, 2023. [Online]. Available https://www.theguardian.com/technology/2023/may/05/geoffrey-hinton-godfather-of-ai-fears-for-humanity.
Hunt, L., & Ferrario, M. A. (2022). A review of how whistleblowing is studied in software engineering, and the implications for research and practice. In *Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software Engineering in Society* (pp. 12–23). Pittsburgh Pennsylvania: ACM. https://doi.org/10.1145/3510458.3513013. May.
Hutchinson, B., et al. (2021). Towards accountability for machine learning datasets: practices from software engineering and infrastructure. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada: ACM* (pp. 560–575). https://doi.org/10.1145/3442188.3445918. Mar.
Johnson, D. G., & Diakopoulos, N. (2021). What to do about deepfakes. *Communications of the ACM, 64*(3), 33–35.
Johnson, S. B., et al. (2022). Cancer misinformation and harmful information on Facebook and other social media: A brief report. *JNCI: Journal of the National Cancer Institute, 114*(7), 1036–1039.
P. Joshi, S. Santy, A. Budhiraja, K. Bali, and M. Choudhury, 'The state and fate of linguistic diversity and inclusion in the NLP world'. arXiv, Jan. 26, 2021. Accessed: May 01, 2024. [Online]. Available: Http://arxiv.org/abs/2004.09095.
Karasti, H. (2014). Infrastructuring in participatory design. In *Proceedings of the 13th Participatory Design Conference on Research Papers - PDC '14* (pp. 141–150). Windhoek, Namibia: ACM Press. https://doi.org/10.1145/2661435.2661450.
Kasirzadeh, A. (2022). Algorithmic fairness and structural injustice: Insights from feminist political philosophy. In *n Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES'22)* (pp. 1–13). Oxford, UK: ACM. https://doi.org/10.48550/ARXIV.2206.00945.
Kelty, C. (2013). Geeks and recursive publics: How the internet and free software make things public. *Emden, christian J./Midgley, david (Hrg.): Beyond habermas. democracy, knowledge, and the public sphere* (pp. 99–118). Oxford/New York: Berghahn Accessed: May 03, 2024. [Online]. Available https://kelty.org/or/papers/unpublishable/Kelty.RecursivePublics-short.pdf.
Khan, M., & Hanna, A. (2022). The subjects and stages of ai dataset development: A framework for dataset accountability. *Ohio St. Tech. LJ, 19*, 171 Accessed: May 01, 2024. [Online]. Available https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/isjlpsoc19&section=9&casa_token=Sa_4CN3Nq7gAAAAA:adtDR7dyqo7moUOBNixkXHKKJZGUpNuZnSkJmpzmUHpXiJWMRmxZrYPPe6_PxqX9p1-s970.
S. Kirchgaessner, 'Israeli spyware company NSO Group placed on US blacklist', *The guardian*, Washington, Nov. 03, 2021. [Online]. Available: https://www.theguardian.com/us-news/2021/nov/03/nso-group-pegasus-spyware-us-blacklist.
S. Kirchgaessner, 'New evidence suggests spyware used to surveil Emirati activist Alaa Al-Siddiq', *The guardian*, London, Sep. 24, 2021. [Online]. Available: https://www.theguardian.com/world/2021/sep/24/new-evidence-suggests-spyware-used-to-surveil-emirati-activist-alaa-al-siddiq.
Klappholz, S. (2024). Open source vulnerabilities dominated 2023, and this year looks no different. *IT Pro, US* [Online]. Available https://www.itpro.com/software/open-source/open-source-vulnerabilities-dominated-2023-and-this-year-looks-no-different.
N. Leveson, 'Medical devices: The therac-25', *Appendix of: Safeware: System safety and computers*, 1995.
P. Lewis and P. Hilder, 'Leaked: Cambridge Analytica blueprint for trump victory', *The guardian*, San Francisco, Mar. 23, 2018. [Online]. Available: https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory.
Lucas, K. T. (2022). Deepfakes and domestic violence: Perpetrating intimate partner abuse using video technology. *Victims & Offenders, 17*(5), 647–659.
McCarthy, J., & Wright, P. (2023). The value of experience centre design to responsible software design and engineering. *Design Issues*.
McCosker, A. (2022). Making sense of deepfakes: Socializing AI and building data literacy on GitHub and YouTube. *New Media & Society*, Article 14614448221093943.
McKeown, M. (2021). Structural injustice. *Philosophy compass, 16*(7), e12757.
Miceli, M., Yang, T., Naudts, L., Schuessler, M., Serbanescu, D., & Hanna, A. (2021). Documenting computer vision datasets: An invitation to reflexive data practices. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada: ACM* (pp. 161–172). https://doi.org/10.1145/3442188.3445880. Mar.
Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2023). Co-shaping an ecosystem for responsible AI: Five types of expectation work in response to a technological frame. *Information Systems Frontiers : A Journal of Research and Innovation, 25*(1), 103–121. https://doi.org/10.1007/s10796-022-10269-2. Feb.
S.G. Monserrate, 'The cloud is material: On the environmental impacts of computation and data storage', 2022, Accessed: May 01, 2024. [Online]. Available: https://mit-serc.pubpub.org/pub/the-cloud-is-material/release/1.
Munn, L. (2023). The uselessness of AI ethics. *AI and ethics, 3*(3), 869–877. https://doi.org/10.1007/s43681-022-00209-w. Aug.

Nabavi, E., & Browne, C. (2023). Leverage zones in Responsible AI: Towards a systems thinking conceptualization. *Humanities and Social Sciences Communications, 10*(1), 1–9 Accessed: May 01, 2024. [Online]. Available https://www.nature.com/articles/s41599-023-01579-0.

Nathan, L. P., Friedman, B., Klasnja, P., Kane, S. K., & Miller, J. K. (2008). Envisioning systemic effects on persons and society throughout interactive system design. In *Proceedings of the 7th ACM conference on Designing interactive systems* (pp. 1–10).

Naughton, J. (2022). Molly Russell was trapped by cruel algorithms of Pinterest and Instagram. *The observer*. Oct. 01Accessed: Jul. 06, 2023[Online]Available https://www.theguardian.com/commentisfree/2022/oct/01/molly-russell-was-trapped-by-the-cruel-algorithms-of-pinterest-and-instagram.

Naughton, J. (2023). There is no moral high ground for Reddit as it seeks to capitalise on user data. *The Observer*. Jun.Accessed: Jun. 17, 2023. [Online]. Available https://www.theguardian.com/commentisfree/2023/jun/17/there-is-no-moral-high-ground-for-reddit-as-it-seeks-to-capitalise-on-user-data.

Newton, O. B., & Stanfill, M. (2020). My NSFW video has partial occlusion: Deepfakes and the technological production of non-consensual pornography. *Porn Studies, 7*(4), 398–414.

Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics, 2*, 25–42.

Nuseibeh, B., & Easterbrook, S. (2000). Requirements engineering: A roadmap. In *Proceedings of the Conference on the Future of Software Engineering* (pp. 35–46).

Pawelec, M. (2022). Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *Digital Society, 1*(2), 19.

Peterson, T. L., Ferreira, R., & Vardi, M. Y. (2023). Abstracted power and responsibility in computer science ethics education. *IEEE Transactions on Technology and Society*.

Purwadi, A., Serfiyani, C. Y., & Serfiyani, C. R. (2022). Legal landscape on national cybersecurity capacity in combating cyberterrorism using deep fake technology in Indonesia. *International Journal of Cyber Criminology, 16*(1), 123–140.

r/software, 'About community'. Accessed: Aug. 04, 2023. [Online]. Available: https://www.reddit.com/r/software/.

I. Rauf, T. Lopez, T. Tun, M. Petre, and B. Nuseibeh, 'Security in online freelance software development: A case for distributed security responsibility'. 2023. Accessed: Jul. 15, 2023. [Online]. Available: https://arxiv.org/abs/2307.06066.

Rowe, N. (2023). It's destroyed me completely:" Kenyan moderators decry toll of training of AI models. *The Guardian*. Aug. 02[Online]. Available https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai.

Sample, I. (2000). What are deepfakes - and how can you spot them? *The Guardian, London*. Jan. 12[Online]. Available https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them.

Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology, 34*(4), 1057–1084. https://doi.org/10.1007/s13347-021-00450-x. Dec.

Scheuerman, M. K., Hanna, A., & Denton, E. (2021). Do datasets have politics? Disciplinary values in computer vision dataset development. *Proc ACM Hum-Comput Interact, 5*(CSCW2), 1–37. https://doi.org/10.1145/3476058. Oct.

Schieferdecker, I. (2020). Responsible software engineering. In S. Goericke (Ed.), *The future of software quality assurance* (pp. 137–146). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-29509-7_11.

Sharp, H., Dittrich, Y., & De Souza, C. R. (2016). The role of ethnographic studies in empirical software engineering. *IEEE Transactions on Software Engineering, 42*(8), 786–804.

Smolka, M., & Böschen, S. (2023). Responsible innovation ecosystem governance: Socio-technical integration research for systems-level capacity building. *Journal of Responsible Innovation, 10*(1), Article 2207937. https://doi.org/10.1080/23299460.2023.2207937. Jan.

Stahl, B. C. (2022). Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence. *International Journal of Information Management, 62*, Article 102441 Accessed: May 01, 2024. [Online]. Available https://www.sciencedirect.com/science/article/pii/S0268401221001341.

Stahl, B. C. (2023). Embedding responsibility in intelligent systems: From AI ethics to responsible AI ecosystems. *Scientific Reports, 13*(1), 7586. https://doi.org/10.1038/s41598-023-34622-w. May.

Star, S. L. (1990). Power, technology and the phenomenology of conventions: On being allergic to onions. *The Sociological Review, 38*(1_suppl), 26–56.

Star, S. L. (2015c). Misplaced concretism and concrete situations: feminism, method, and information technology. *Boundary objects and beyond: Working with leigh star* (pp. 143–167).

Star, S. L. (2015b). Revisiting ecologies of knowledge: Work and politics in science and technology. *Boundary objects and beyond: Working with leigh star* (pp. 13–47). Cambridge, MA, USA: MIT Press.

Star, S. L. (2015a). The ethnography of infrastructure. *Boundary objects and beyond: Working with susan leigh star* (pp. 473–488). Cambridge, MA, USA: MIT Press.

Star, S. L., & Bowker, G. C. (2006). How to infrastructure. *Handbook of new media: Social shaping and social consequences of ICTs* (pp. 230–245).

Star, S. L., & Ruhleder, K. (1994). Steps towards an ecology of infrastructure: Complex problems in design and access for large-scale collaborative systems. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work* (pp. 253–264).

Star, S. L., & Ruhleder, K. (2016). 20 Steps toward an ecology of infrastructure: Design and access for large information spaces. *Boundary Objects and Beyond: Working with Leigh Star*, 377.

Stilgoe, J., Owen, R., & MacNaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy, 42*, 1568–1580.

Synopsys Inc, 'Mew synopsys report finds 74% of codebases contained high-risk open source vulnerabilities, surging 54% since last year'. Accessed: May 03, 2024. [Online]. Available: https://www.prnewswire.com/news-releases/new-synopsys-report-finds-74-of-codebases-contained-high-risk-open-source-vulnerabilities-surging-54-since-last-year-302071630.html.

Thimbleby, H. (2021). *Fix IT: See and solve the problems of digital healthcare*. Oxford: Oxford University Press.

Université de Montréal, 'Montréal declaration responsible AI', Montréal Declaration of Responsible AI, Montréal. [Online]. 2018. Available: https://recherche.umontreal.ca/english/strategic-initiatives/montreal-declaration-for-a-responsible-ai/.

S. Vaughan-Nichols, 'Ethical-source movement opens new open-source organization'. Accessed: Jan. 19, 2023. [Online]. Available: https://www.zdnet.com/article/ethical-source-movement-opens-new-open-source-organization/.

Veinot, T. C., Mitchell, H., & Ancker, J. S. (2018). Good intentions are not enough: How informatics interventions can worsen inequality. *Journal of the American Medical Informatics Association, 25*(8), 1080–1088 Accessed: May 03, 2024. [Online]. Available https://academic.oup.com/jamia/article-abstract/25/8/1080/4996916.

Walker, A. (2022). Molly Russel Inquest - Coroner's conclusion in full. *The Independent, London*. Sep. 30Accessed: May 19, 2023. [Online]. Available https://www.independent.co.uk/news/uk/molly-russell-inquest-coroner-s-conclusion-in-full-b2183287.html.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review, 9*(11).

Widder, D. G., Nafus, D., Dabbish, L., & Herbsleb, J. (2022). Limits and possibilities for "Ethical AI" in open source: A study of Deepfakes. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2035–2046).

Wikipedia, 'GNU General Public Licence'. Accessed: May 19, 2023. [Online]. Available: https://en.wikipedia.org/wiki/GNU_General_Public_License#:~:text=The%20GNU%20General%20Public%20License,share%2C%20and%20modify%20the%20software.

Williamson, B., Macgilchrist, F., & Potter, J. (2023). *Re*-examining AI, automation and datafication in education. *Learning, Media and Technology, 48*(1), 1–5.

Wilson, T. D., Gordon, W. T., Lipson, A. W., & Thavarajah, B. M. (2020). Deepfakes" pose significant market risks for public companies: How will you respond? *The Journal of Robotics, Artificial Intelligence & Law, 3*.

Winter, R., & Salter, A. (2020). DeepFakes: Uncovering hardcore open source on GitHub. *Porn Studies, 7*(4), 382–397.

Young, I. M. (2006). Responsibility and global justice: A social connection model. *SOY, 23*(01), 102. https://doi.org/10.1017/S0265052506060043

Young, I. M. (2010). Responsibility and global labor justice. *Responsibility in Context: Perspectives*, 53–76.

Young, I. M. (2011). Responsibility for justice. *Oxford political philosophy*. Oxford: Oxford University Press.

Zong, J., & Matias, J. N. (2024). Data refusal from below: A framework for understanding, evaluating, and envisioning refusal as design. *ACM J Responsib Comput, 1*(1), 1–23. https://doi.org/10.1145/3630107. Mar.