NOTE: this is a pre-publication version. There may be some changes with the final published version.

Please cite as: Thompson, T. L., & Graham, B. A more-than-human approach to researching AI at work: Alternative narratives for human and AI co-workers. In Bonderup Dohn N, Jørgen Hansen J, Børsen Hansen S, Ryberg T & de Laat M (eds.) *Conceptualizing and innovating education and work with networked learning*. Springer, pp. 171-187. https://doi.org/10.1007/978-3-030-85241-2_10.

A more-than-human approach to researching AI at work: Alternative narratives for human and AI co-workers

Terrie Lynn Thompson
Faculty of Social Sciences, University of Stirling, terrielynn.thompson@stir.ac.uk

Bruce Graham
Computing Science and Mathematics, Faculty of Natural Sciences, University of Stirling, bruce.graham@stir.ac.uk

Abstract
Professional workers practice at the intersection of public narratives about artificial intelligence (AI), the AI industry, and regulatory frameworks. Yet, there is limited understanding of the interactions between workers, AI systems, and the publics they serve. To inform networked learning scholarship, there is a pressing need to study the knowledge that workers are developing as they learn to work with AI and the implications for networked learning within the workplace and higher education. We bring social and computing science perspectives alongside more-than-human sensitivities to explore how professional expertise, judgment, accountability, and control is being re-distributed between human workers and AI systems. By sketching the changes AI is provoking we highlight the fine-grained research and analysis necessary to ensure that AI design and deployment is critically informed by in-depth understandings of how people are actually engaging with algorithmic systems. We raise questions about what trust and confidence in new AI-infused work practices is needed (or possible). Attention is drawn to the complexities of AI-mediated work, which invites re-thinking ways to generate the evidence needed to inform networked work-learning practices. Highlighted throughout is the power of AI narratives and the importance of advancing alternative, more nuanced, narratives.

Keywords
artificial intelligence, networked learning, professional work, ethics of technology, more-than-human, public understanding of technology

Introduction

As artificial intelligence (AI) weaves its way into everyday work, learning and living, labour is being re-distributed between workers and their new digital counterparts. Globally, national policies present ambitious aspirations for rapid uptake of AI, positioned as a key driver of innovation, labour productivity, and economic growth that needs to be advanced swiftly in order to attain global competitiveness and leadership. AI is also seen as key to finding solutions for critical societal challenges including the UN Sustainable Development Goals.

However, it is not clear what impact AI has, and should have, on workers, particularly professional workers; what work-related policies and organizational practices are needed to address these changes; or the learning implications for professional workers as they interact more intensely with various forms of AI. Largely thought to be immune from automation, professional work is now challenged as AI increasingly adds advanced data analytics to augment complex professional decisions, automates tasks, and enables new forms of remote working (e.g., Susskind & Susskind, 2015).

To better inform networked learning scholarship, there is a pressing need to understand: (1) the new competencies and knowledges workers are developing as they learn to work with AI; and (2) implications for professional learning within the workplace and higher education (HE). Our aim is to contribute to this conversation by sketching some of the changes AI is provoking in workers' day-to-day practices in order to highlight the fine-grained research and analysis needed to ensure that AI design and deployment is critically informed by an in-depth understanding of how people experience and engage with such algorithmic systems.

Following other approaches (European Commission (EC), 2019; Nilsson, 2010), our working definition of AI is any computational system that carries out a task normally associated with a degree of intelligence when performed by humans. The rising prominence of complex AI systems in the workplace is challenging professional roles and skills as new decision-making processes distribute judgment and responsibility across AI-human systems. Coming to the fore is the trustworthiness of AI outputs, as emphasized in policy recommendations by The High-Level Expert Group on Artificial Intelligence of the European Commission (2019).

Increased use and trust in AI to deliver professional services depends on an informed, critical, and willing public. However, the escalating debate about the incursion of AI into the workplace remains stubbornly polarized. Recent reports attempting to gauge public perception suggest that amidst exaggerated expectations and fears about AI, citizens are sceptical, believe "it won't happen to me", and lack understanding of what AI is and does (Archer et al., 2018). Others point to the divergence between the AI hype and the views of experts (e.g., Bristows, 2018).

AI narratives have long been influenced by fiction, which fan the fear of robots replacing humans and depict versions of AI that are well beyond current or even near future reality. These narratives are important (The Royal Society (RS), 2018). Critically informed and positive, they drive ethical investment and innovation at all stages of development from research to commercialization as well as robust AI-related education and learning initiatives that foster effective worker-machine interactions. However, negative perceptions, and especially those fuelled by spurious narratives, could lead to public backlash that curtails AI development and further entrenches misinformation and distrust.

Professional workers practice at the intersection of prevailing narratives about AI, professional regulatory frameworks, the fast-paced AI industry, and their own competencies and degree of trust regarding AI systems. We take a broad view of the professional worker: a member of an occupational group "that defines itself as collectively sharing particular knowledges and practices, and that is publicly accountable for its service" (Fenwick & Nerland, 2014, p. 2). Although the impact of AI on work is far-reaching, much of the current focus is on macro-employment trends: jobs gained/lost, what work can be automated, and re-skilling the workforce for the "jobs of tomorrow". Missing in these narratives is an understanding of the significant changes in *work itself* and the learning opportunities inherent in these new arrangements of work as AI becomes more prominent.

Because work and work-related learning are often inextricably linked, looking at both enables deeper understanding AI and workers' networked ways of learning *and* working. Focusing on examples from several sectors (including HE), we begin by exploring how professional expertise, judgment, accountability, and control is re-distributed as workers interact with AI systems. Evident in these examples are themes that resonate across different sectors and forms of professional work. We raise questions about what trust and confidence in new AI-infused work practices is needed (or possible) and how is this negotiated. In so doing we draw attention to the complexities of AI-mediated work practices, which invites re-thinking ways to research these shifts in order to generate the evidence needed to inform networked learning practices. Given the early stages of this field of inquiry, we hope to evoke discussion of alternative human-AI narratives suited for the messy—and often unseen—realities of everyday practices and consider the implications for researching these practices.

In this paper we make three contributions to networked learning research. First, we situate networked learning more tightly with work itself. Learning emerges in relation to specific tasks, technologies, and responsibilities: activities and goals in a workplace structure the curriculum of the workplace (Ludvigsen & Nerland, 2018). Attending to the "pedagogy of work"—which we refer to as *networked work-learning*—highlights ways to exploit these learning opportunities and identify emerging competencies. Second, we align with the reassertion of the political and moral aspirations of networked learning to help people build necessary capabilities for constructing better ways of living (NLEC, 2020). We therefore focus on how it might be possible to live and work critically, ethically, and productively *with* AI in order to push against reductionist, deterministic, and instrumental conceptions of human-technology endeavours. Third, we build on views of learning as connecting: people; sites of learning and action; ideas, resources and solutions; and across time, space and media (NLEC, 2020, p. 8). We extend the conceptual framing of connectivity within networked learning by engaging in further conversation with more-than-human sensibilities. Rather than human vs. technology, this perspective attunes to the human-technology together as the phenomena of interest. In so doing, the focus is on how changes to networked work-learning are produced by a series of complex social and material (digital) relations. Such theorization may offer insights into research needed to untangle the interweaving of AI and work-learning.

Negotiating with AI: Re-distribution of Professional Work-Learning

The rapid pace of recent AI advances is driven by machine-learning algorithms, including deep learning; exponential increase in computing capacity which can train larger and more complex models much faster; and vast amounts of data (Manyika et al., 2017). Such shifts are shaping assertions that "we are on the cusp of a new automation age in which technologies not only do things we thought only humans could do, but can increasingly do them at a superhuman level" (Manyika et al., 2017, p. 24). However, current discourse on AI and its impact on professional services suggest that AI debate and research is in the early stages and does not yet untangle important distinctions and complexities. Necessary to inform next steps in AI-related development and policy is an understanding of the significant changes in work itself and the learning opportunities inherent in arrangements of work.

AI can do a range of tasks with varying degrees of sophistication: intelligent decision support, classification, prediction, visual object recognition and image processing, speech recognition, natural language processing, and natural language generation. AI is not one thing, and often invisible, resulting in complex changes to work not yet well understood.

There is limited evidence of how AI is being used currently and how workers' tasks have changed where this has happened (Frontier Economics, 2018). Professional bodies responsible for profession-specific regulations and codes of conduct are grappling with drastically changing professional work landscapes, ethical dilemmas, and a desire to seize opportunities afforded by AI while also minimizing risk.

A number of questions are posed. Edwards and Fenwick (2016) ask how we think about professional responsibility and accountability when decisions are delegated to complex digital systems or what it means to consider a professional as a responsible agent when capability is distributed across human and digital actors. Evidence is needed of how AI-mediated work practices are changing decision-making processes, the valuing of professional judgment, and newly distributed responsibilities for algorithmic-influenced decisions. These sorts of research questions are relevant to HE: not only as a sector and workplace but also as the space where future workers should be learning how to negotiate and critically question complex, fast changing, digitally mediated ways of working.

Allert and Richter (2018) highlight a profound shift: as automation and algorithmitization of knowledge work turn data into a resource for, and product of, computation, certain regimes of knowledge that replace subjective experience with objectified data come to the fore. In addition to delegating routine tasks to AI, complex decisions are increasingly based on computational analysis of big data raising questions about the capacity and need for human judgment. Although decision makers may be reluctant to depart from algorithmic recommendations (thus further undermining individual judgment and discretion), others argue that not all decisions can be coded (Agrawal et al., 2019). Indeed, the phenomenon of "algorithm aversion" suggests that humans are less confident about accepting and using the results of automated forecasting and prediction (Dietvorst et al., 2015).

As professionals undertake new and different responsibilities for knowledge, understandings of where "expert" knowledge resides becomes blurred. Lange et al.'s (2019) study of algorithms and high-frequency trading suggests that most of the time "neither the human trader nor the algorithmic machine is in full control", highlighting the constant reconfiguration of worker-algorithmic relations (p. 600). The outsourcing of work activities to, and with, algorithms is leading to new forms of "algorithmic management": prolific data collection and surveillance, transfer of performance evaluations to rating systems or other metrics, and the use of "nudges" and penalties to indirectly incentivize worker behaviors (Kolbjørnsrud et al., 2016; Mateescu & Nguyen, 2019). In a study of AI-mediated decision-making in a telecommunications company, Bader and Kaiser (2019) document how the user interface mediates between human and algorithmic decisions. Significantly, they note that a lack of balanced involvement of humans in decisions has negative performative effects due to deferred decisions, workarounds, and manipulations (Bader & Kaiser, 2019). Clearly, workers are now navigating very different social and material relations, presenting significant implications for professional learning within the workplace and higher education.

In the next section we highlight the complexities of these shifts in responsibility and control by drawing on examples from several sectors reflective of current narratives. Foregrounded in these examples are tensions in the openness of AI systems and the data upon which they build; increasing reliance on public-private technology partnerships; contradictory rhetoric about AI and its actual level of uptake in provision of professional services; bias in datasets used for training AI; calls for holding algorithms accountable despite often messy human-AI partnerships; and the need for co-development of algorithms and AI systems.

New Dilemmas of Professional Work

As reported by Tromans (2019), the recent ban obtained by France's judges on the use of public court data for the statistical analysis and prediction of their decisions in court (i.e., legal predictive analytics) has led the French National Bar Council to demand that lawyers should also be excluded from statistical analysis of their actions in court. France may be the "first country in the world where litigation analysis and predictive modelling face such a comprehensive ban" (para 6). In light of France's "Open Data" movement, intended to make all public data available online, Tromans (2019) points to contradictions in the emergence of a "two-tier" public data system: "citizens can know some things, but not others, even when the underlying information is public" (para 13). Moreover, the work of legal professionals and court practices are further obscured with some lawyers claiming this move as "irreconcilable with their mission to represent and defend their clients" (para 15). The tensions evident in the French court system relate to the openness of AI systems and the data upon which they build.

Further concerns arise when AI systems move from merely informing to prescribing professional decisions and actions. In the case reported by the *AI Now Institute*, the use of student test data to make teacher employment decisions including promotions and terminations revealed (in a subsequent lawsuit) that no one in the school district could explain or replicate the determinations made by the system even though the district had access to all the underlying data (Whittaker et al., 2018). The teachers who contested the AI outputs were told that the system was simply to be believed and could not be questioned. After the vendor fought against providing access to detailed information on how its system worked, and a ruling that such an AI system could contravene constitutional due process protections, the school district eventually abandoned the third-party AI system in question.

Private-public partnerships that often sustain extensive use of AI systems in the provision of professional service are potentially problematic as decision-making, responsibility, accountability, and the underlying data are not only increasingly distributed across a range of actors but sometimes "black boxed". Predictive algorithms can be used in criminal justice systems to inform decision-making in policing patterns as well as bail and sentencing decisions. Described in a recent Council of Europe (2018) motion as effective systems valued by the authorities that use them, they nevertheless urge attention to: (1) how such systems are usually provided by private companies and not subject to public scrutiny; and (2) how police departments may lose control over their own data and become dependent on the private companies that have acquired this data.

Let's consider automation, another area of lively debate. Chatbots are one form of automation increasing in popularity and sophistication. Smutny and Schreiberova (2020) describe chatbots as software that interacts with users "in a natural, conversational way using text and voice" (p. 1). Deployment of the Jill Watson chatbot, as a teaching assistant on AI courses at Georgia Institute of Technology, was intended to provide co-teaching support. Indeed, Eicher et al. (2018) note the importance of humans working alongside the AI system and stepping in, for example, when the chatbot could not handle uncommon cases. However, a significant amount of data and expertise is required to create the conversational flow that people have come to expect as they interact with personal digital assistants (Smutny & Schreiberova, 2020). Such development demands extensive collaboration between computing/data scientists and educators. Although somewhat blackboxed by the chatbot, it is possible to see how "teaching" expertise is *distributed* across spaces and time: it is both automated and not. Here we see an important but subtle shift in the rhetoric of automation of professional workers. As Bayne (2015) suggests, teacher automation does not need to be about replacing or "solving" productivity deficits in teachers but rather, can take on a more distributed conception of teaching work that considers how assemblages of "teacher-student-code might be pedagogically generative" (p. 465).

Adding to the complexity of AI are the contradictions in the current rhetoric about AI and its actual uptake in professional work and services. This is evident in an ethnographic study on the use of AI-mediated risk-assessment tools in the USA criminal court system. Christin's (2017) analysis suggests that such AI systems are often actively resisted in criminal courts and are far less powerful and persuasive than suggested in the current narratives extolling widespread AI deployment. She notes that because the judges and prosecutors in her study did not trust the algorithms (they did not know the companies they come from, they did not understand their methods, and often found them useless), the AI outputs often went unused (para 12). Christin (2017) describes how the software was used, score sheets printed out and added to the defendants' files, after which the "scores then seemed to disappear and were rarely mentioned during hearings" (para 12). Foregrounded is the importance of attending to everyday practices. Interestingly, Christin (2017) found that the issue creating resistance was not the transition to complex AI risk-assessment tools per se but rather the more basic transition to paperless case-management systems.

Despite slow uptake of AI in education, AI technology is increasing the range of applications possible within learning analytics (LA) systems. LA systems are co-workers of sorts, helping instructors analyse student behaviour and performance with some LA automatically tailoring teaching material to individual students. In this way, the expertise and judgement of teaching is now a shared responsibility and distributed across space and time. Amidst troubling narratives of a more pernicious data gaze (aka Beer, 2019) are questions of how this gaze *by* and *on* professional workers amplifies both visibility and invisibility. Brown's (2020) study of five HE physics instructors who used data dashboards (designed to deliver "algorithmically assembled information about students to the instructor") suggests that LA systems can be employed by institutions to surveil faculty as well as students (p. 388). Brown (2020) reports that the dashboards facilitated data collection about instructors' pedagogical planning and decision-making that threatened their sense of autonomy, opened for the door for unwarranted interference, undermined their existing pedagogical strategies, and enabled unwelcome surveillance. Here, increased digital visibility of work shines a light on a darker side of algorithmic management of professional work.

Trusting AI Co-Workers

From the outset, networked learning has been concerned about the learning possibilities enacted through connections between people, technologies, ideas, resources, and contexts. The examples in the previous section start to raise questions about the knowledge and knowing practices workers are developing as they work with AI, signalling potential areas for research. Foregrounded is the challenge of how workers come to trust (or not) their AI co-workers; something that unfolds through complex connections between people, technologies, professional expertise and judgement. The EC (2019) identifies trustworthy AI as a foundational ambition: not only the technology's inherent properties, but adopting a socio-technical approach that attends to both human and technology actors throughout the AI ecosystem and life cycle. There are good reasons for caution. Bias and lack of transparency in how algorithms work are shortcomings in current AI systems and an active area of research.

Addressing these issues is crucial for developing AI systems that workers and the public trust and want to work, learn, and live with. If we do not want blackboxed technologies, Bunz (2017) argues that it is essential to learn how to interact with them more attentively (p. 253). Without this attentiveness, there will be repercussions. For example, consider Uber's deliberate obscuring of the algorithms that determine demand and supply pricing of fares, which led to drivers to "game the system" in order to control and create price surges (Rosenblat & Stark, 2015). The efforts of these workers to address information asymmetries highlight the consequences of imposed algorithms that are not transparent or trusted by workers. One might image similar scenarios playing out in HE by students and staff subjected to similarly blackboxed but nevertheless power-imbued algorithms. Successful deployment and use of AI in the workplace relies on human's acceptance of and trust in their AI co-workers. Underscored is the importance of understanding the unease towards AI—which includes asking the tough critical questions—and then working to address concerns.

While there is considerable evidence for data-driven algorithms outperforming human experts across a wide variety of domains, it seems humans are less forgiving of errors made by algorithms than by humans (Dietvorst et al., 2015). Better understanding this predilection is important for humans and AI to learn how to work together. Yu et al.'s (2018) research of human-algorithm interaction suggests that during ongoing interaction with an algorithm, humans will assess the apparent reliability of the algorithm and adjust their acceptance of its outcomes accordingly: in this way, "acceptance thresholds are dynamic and user-dependent" (p. 262). Dietvorst et al. (2018) suggest that allowing human users a degree of control, such as the ability to modify the algorithm, may ease some of the tensions between workers and the AI systems with which they interact.

In the light of evidence of human mistrust in AI and algorithmic systems, research is required to understand how an AI system could be configured to become a trusted entity within a mixed human-AI working environment. From a networked work-learning perspective, this entails examining the constantly shifting connections between humans and the myriad of digital actors that comprise AI systems. For example, Robb et al. (2018) identify several factors that impact user confidence: trade-offs between abstraction and detail in the presentation of *algorithm outputs* to different types of user (naïve versus expert); how much explanation of *algorithm operation* is required (again, may be user dependent); need for information on *data provenance* (for data-driven and trained algorithms).

However, it is not merely the functionality of AI at issue. The examples illustrate how it is both AI and humans together that enact professional work. Given that many of the current AI narratives set up an ontologically distanced relationship between these complex digital assemblages and human actors, we argue that a more co-constitutive perspective helps to avoid over-simplistic deterministic stances. As Kitchin (2014) states, AI does not "exist independently of the ideas, instruments, practices, contexts and knowledges used to generate, process and analyse it" (p. 2). The need for such sensibilities is highlighted in this next example.

Failure to appreciate the complex material and social environments into which AI systems are enrolled can lead to high-profile disasters, such as the decision to use a computational algorithm to rebalance grades given by teachers (based on coursework) in the wake of cancellation of school exams in the UK in spring 2020 due to the COVID-19 pandemic. Both England and Scotland introduced hand-crafted algorithms based on current and historic performance data across schools and student cohorts (Ofqual, 2020; Priestley et al., 2020) with the aim of ensuring the 2020 grades for English A-levels and Scottish Highers would be in line with past performance (Bedingfield, 2020). However, what unfolded was a reduction of the teacher-predicted grades for many students, often to below that needed for university entrance. In particular, high-performing students at otherwise poorly performing schools were hard hit. The resulting uproar led to the abandonment of the use of these algorithms. A report commissioned by the Scottish Government acknowledged the difficult and time-poor circumstances in which the model was developed and deployed but highlighted issues with the inequities and lack of transparency in the algorithms; the way emergent events amplified the uncertainty of decision-making; lack of communication and engagement with teachers and parents around the process and algorithm; the perceived arbitrary nature of the approach; and perhaps most disappointing, an erosion of confidence in the Scottish Qualification Authority and damaged relations between some students and their teachers (Priestley et al., 2020). Bedingfield (2020) sums up: "the algorithm has been ditched, and students will be belatedly graded with the original teacher's predictions" (para 3). A stark reminder of how teachers' expertise and judgement is necessary but also must be necessarily re-distributed in thoughtful ways.

This example highlights the difficulties in developing a sophisticated and robust algorithm for complex predictive or decision-making scenarios and, more importantly, how to deploy an algorithm in a way that contributes positively to the work of the professionals using it and the people affected by its outcomes. The lack of transparency in the process was compounded by the lack of involvement of teachers and education specialists in key decisions. Questions remain about how the professional expertise and judgement of teachers in this situation was viewed and performed. Although government was challenged about decisions and processes, the algorithm itself was widely criticized with headlines such as "Ditch the Algorithm" (Amoore, 2020 writing in The Guardian). Although many narratives became polarized around the algorithm, the algorithm did not act alone. It takes humans, technologies, and a range of actors to co-constitute these new forms of work. Professional agency, expertise, judgement, and accountability: these are assemblages of algorithms, interfaces, data, teacher-student-parent relations, educational specialists, statisticians and data scientists, statistical models, and policy.

Quite rightly, Amoore (2020) argues that this type of decision-making involves far more than a series of computational steps. She states that "grappling openly and transparently with difficult questions, such as how to achieve fairness, is precisely what characterises ethical decision-making in a society" and technical questions about data inputs and weighting of features are "political propositions about what a society can and should be like" (Amoore, 2020 para 9/6). Here is an example of an alternative AI narrative. Foregrounded is the importance of connections between learning and the kind of change that it is considered important in the world (NLEC, 2020). Indeed, innovative networked work-learning research may help to navigate—conceptually, ethically, and practically—these fluid social and material relations to better understand and approach the re/dis-assembling of AI-human entanglements.

Making AI Visible

Adding to the challenge of understanding how trust develops is that AI is often invisible, making it difficult for people to understand how and when they interact with it (Bristows, 2018). The problem is exacerbated by the increasing availability of (relatively) easy-to-use software tools for creating data-trained AI systems (e.g., deep neural networks). Some AI systems can now be built by people who have little understanding of the inner workings of such systems and their limitations.

Argued is the need for explainable AI (XAI), seen as essential if workers are to "understand, appropriately trust, and effectively manage an emerging generation of AI systems" (Gunning & Aha, 2019, p. 45) and is meant to afford humans a degree of functional understanding of AI outputs. If people do not know how AI arrives at decisions, they will not trust it; an issue attributed to the failure of IBM Watson for Oncology, an AI system designed to assist doctors with cancer diagnoses. Polonski (2018) highlights the tensions that emerged in the deployment of IBM's AI system:

> If Watson provided guidance about a treatment that coincided with their own opinions, physicians did not see much value in Watson's recommendations. The supercomputer was simply telling them what they already know, and these recommendations did not change the actual treatment. … [If] Watson generated a recommendation that contradicted the experts' opinion, doctors would typically conclude that Watson wasn't competent. And the machine wouldn't be able to explain why its treatment was plausible because its machine learning algorithms were simply too complex to be fully understood by humans. Consequently, this has caused even more mistrust and disbelief, leading many doctors to ignore the seemingly outlandish AI recommendations and stick to their own expertise. (paras 5-6)

However, an important question arises about worker and public expectations of an AI system: Is the expectation to replicate human expertise and/or to improve upon it? If it is the former, then we would likely expect to be able to interrogate the AI to understand how it has arrived at an output, in the same way we could ask a human expert. That said, if we can accept that the AI system may work differently from human reasoning and potentially with higher performance, could workers and the broader publics accept that a human-understandable explanation of how the AI works may not be possible?

The operation of many AI technologies—rule-based systems, case-based reasoning, decision trees—is transparent to humans. An approach to XAI is to try to use these technologies to model the performance of non-transparent AI systems, such as deep neural networks (Ribeiro et al., 2016). The downside is that any "explanation" that arises is still only an approximation to what the AI is really doing. That said, the same may be true for a human expert asked to give an explanation of how they reached a conclusion.

In this situation, the important factor in deploying AI in the workplace is whether adding such a level of explanation provides increased and necessary trust in the AI; even if the explanation is not strictly accurate (Robb et al., 2018). Truly powerful AI systems may not be understandable and therefore the entire AI ecosystem (which includes designers, industry, policy makers, workers, researchers, and the public) needs to find other ways of establishing trust in such systems. This could include continual monitoring of the utility of the outcomes produced by the AI so that trust is established via increasing confidence in the robustness and performance of the AI. We suggest that deployed AI systems should come under critical performance appraisal in the same way as a human employee.

One challenge to the development of trustworthy AI is built-in bias. Because humans exhibit bias in decision-making either consciously or unconsciously, a potential selling point for AI decision support systems is their lack of bias. Unfortunately, this is difficult to achieve in practice, as it requires large and truly representative data sets to underpin the training of the AI. For example, Hao (2019) explains how risk assessment tools used in the justice system are designed to generate a recidivism score (a single number estimating the likelihood that a person will reoffend) that is then used by a judge to help determine what type of rehabilitation services particular defendants should receive. However, Hao (2019) points out that such tools are often driven by algorithms trained on historical crime data, which means that populations that are historically disproportionately targeted by law enforcement (e.g., low-income and minority communities) are at risk of high recidivism scores. These algorithms may in fact “amplify and perpetuate embedded biases and generate even more bias-tainted data to feed a vicious cycle” (Hao, 2019, para 10).

Eicher et al. (2018) explain that the task of giving a personal response to every student introduction was delegated to Jill Watson, the chatbot employed at Georgia Tech. However, they “realized that while the system’s training made the chatbot capable of responding to a phrase like ‘will be father for the first time’ … it would not react specifically to something like ‘I’m pregnant’" (p. 90). It was at this point that Eicher et al. (2018) realized they were dealing with biased data sets. In creating Jill Watson, they used the common practice of building responses based on posts from previous semesters, rather than trying to speculate about what a student might post. Given that women are a minority group in their computer science program one can see how a data set of previous postings could be gender biased. Eicher et al. (2018) comment: “We are particularly sensitive to this issue … so much effort is going into providing a more welcoming environment for minorities. We now actively monitor the selection of answers and comments she’s [Jill Watson] capable of offering to detect and correct any signs of such bias” (p. 93).

The issue of bias in datasets and algorithms is widely recognised by AI developers and is a necessary part of the public AI narrative on the limitations of AI systems. The onus is now on the *range of actors* involved in the AI ecosystem to understand and to identify—in practice—the limitations and biases of the system and to work towards generating genuinely unbiased—trustworthy—datasets for use in training AI.

An Uneasy Alliance

Perhaps the best way to describe the current situation is an uneasy alliance: there are many aspects of work that can be done better and in ways that do not minimize, devalue, or exclude the human. But there are also many potential uncertainties and dilemmas. It is possible to build on the opportunities created by the current wave of AI systems. Polonski (2018) provides examples of how police forces use AI to map when and where crime is likely to occur and how doctors can use it to predict when a patient is most likely to have a heart attack or stroke. There is evidence of significant economic benefits when AI is used to optimize production processes, especially when coupled with suitable workforce retraining in the AI technologies to avoid staff layoffs (Partnership on AI (PAI), 2018). Image processing by deep neural networks (Le Cun et al., 2015) is a strong success story for AI, promising to cope with examining large volumes of medical imaging data for signs disease, and even able to find disease indicators in such data that are not evident to human experts.

AI developed in-house by Zymergen, a start-up company in the USA to automate laboratory services, found that close collaboration with laboratory scientists during the AI development was crucial to establishing trust in the end systems (PAI, 2018). Such collaboration between AI developers and workers is extremely important. Deepening involvement with AI systems not only distributes, but also amplifies, workers’ implicatedness (Thiele, 2014) and thus expands their ethical responsibilities. Workers therefore need to be part of the design and development of responsible human-AI interaction in ways that do not minimize human intelligence or capabilities. AI narratives are beginning to reflect the need for increased co-development involving educators, AI technologists, data scientists, workers, and the various publics they serve (e.g., Luckin & Cukurova, 2019).

We have drawn on descriptive accounts to foreground some of the complexities of AI-mediated work and how expertise, judgement, control, trust, and accountability is being re-distributed as workers work and learn with AI systems and intensified datafication practices. However, generating research evidence to inform networked work-learning scholarship, demands innovative approaches to attune to the narratives, nuances and often hidden, yet situated, interactions between humans and algorithms. This is where we turn next.

Researching with More-than-Human Sensibilities

In this article we draw on a more-than-human orientation to untease and describe connections between people, technologies, ideas, resources, and contexts. These connections are of interest to networked learning research. Much of the current discourse around AI systems reinforces the binaries of human vs. machine, worker vs. AI, and human vs. artificial intelligence. Workers and AI systems are often portrayed as somehow connected, but separate, entities. And yet, many current and promising uses—and robust critical questioning—of AI systems in the provision of professional services seem to be about how AI systems and humans work together.

We suggest that more-than-human sensibilities provide a way to conceptualize, attune to, and study the complex interactions that unfold between AI systems, workers, ways of working, workplaces, policies, and public narratives. Networked work-learning practices are seen as distributed across multiple networks and changes to work and professionalism as a series of complex social and material (digital) relations. AI systems introduce a myriad of new actors and connections into these networks—in many instances, new assemblages emerge. Understanding the larger shifts and the ethical implications demands sensibilities, theory, and methodologies to see the human-technology together as the phenomena of interest. This leads to more-than-human accounts that offer a new and more inclusive account of what it means to be human in an increasingly technologized world.

Such conceptual and methodological work is an important contribution to be made by social sciences. It is beyond the scope of this article to articulate the many ways such research might be undertaken. Our focus has been to outline potential areas for research and possibilities for conceptualizing new questions. Bucher’s (2016) technographic approach and heuristics for interviewing objects (Adams & Thompson, 2016) open up possibilities. Throughout we have pointed to how prevailing AI narratives are powerful actors and the importance of exploring and advancing alternative, more nuanced, narratives. One challenge is the small number of similar and potentially misleading narratives that dominate public debate. Indeed, prevailing AI narratives are seen to contribute to some of the mistrust and unease about these powerful technologies (Lemay et al., 2020). The narratives about AI prevalent in public discourse inevitably shape the deployment of AI in the workplace as well as the kinds of research questions that are considered important to study. Research focused on examining AI-human interactions in work and learning needs to take these narratives into account.

Changing AI Narratives: Re-Assembling Actors

Public perception of AI is shaped by hundreds of years of stories that people have told about humans and machines, often of a dystopian nature. In these stories, AI is embodied (a robot) and super-intelligent, a trope that leads to inflated expectations and fears about the technology and influences the way the technology is portrayed in popular culture and the media. It is important to recognize that AI deployment in various work sectors is currently performed in the context of workers and publics who bring expectations and beliefs about

AI: accurate or not. A recent report by The Royal Society (RS, 2018) summarises the common narratives and their drivers. As an easy target for sensationalism and hype, stories about AI often reinforce fears and/or hope for its future potential of AI and muddy the waters as to its immediate possibilities (e.g., if and when the "AI singularity" will happen). Understanding, acknowledging, and then pedagogically addressing these perceptions in order to clarify and educate workers and the publics they serve about the realistic possibilities of AI in the provision of professional services is vital to successful deployment.

More-than-human sensibilities align with methods such as controversy mapping (e.g., Venturini, 2010) or networked ethnography (Ball et al., 2017). These approaches enable the researcher to examine and articulate how narratives, and therefore, knowledge about AI emerges and moves via complex social, technical and political constellations of actors, texts, and technologies as a form of assemblage. There is also a place to utilize innovative participatory research methods to enable new AI narratives to emerge through two-way public dialogues. This is consistent with the ethos of networked learning and its stronger focus on inquiry and action. One possible approach are *mini-publics* (Escobar & Elstub, 2017): assemblies of citizens brought together to learn and deliberate about the use of AI systems and the impact this has on confidence in the changing provision of professional services in order to inform public opinion and decision-making. Mini-publics create spaces for the public to learn about and experience AI first-hand, actively shape the direction in which technology progresses, and interact directly with social science and computer/data science experts. Work by the RS (2018) illustrates how it may be possible to re-craft compelling narratives about AI that accurately reflect "the underlying science and its possibilities while acknowledging scientific and social uncertainties" (p. 20).

Our aim is to spark discussion about new research directions that engage with alternative human-AI narratives suited for the messy—and often unseen—realities of everyday AI-mediated professional work practices. There is a substantial role for networked learning research and practice in this space. Indeed, the *Automating Society* report (Chiusi et al., 2020) emphasizes the pressing need to enhance algorithmic literacy and strengthen public debate on automated decision making systems and has included this as a key policy recommendation to the EU parliament. There are opportunities to enable people to work with data-enabled, AI-powered systems in ways that give them a better understanding of their collective entanglements with AI and networked work-learning practices. In so doing, it may be possible for human workers to critically know their AI co-workers, in the same way they know human colleagues—their strengths, weaknesses and biases—and visa versa.

References

Adams C., & Thompson, T.L. (2016) *Researching a Posthuman World: Interviews with Digital Objects*. Palgrave Macmillan. 10.1057/978-1-137-57162-5

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Exploring the impact of artificial Intelligence: Prediction versus judgment. *Information Economics and Policy, 47*, 1–6.

Allert, H., & Richter, C. (2018). Perspectives on data and practices. *Proceedings of the 10th ACM Conference on Web Science,* 173-176. ISBN: 978-1-4503-5563-6. ACM.

Amoore, L. (2020, Aug 19). Why 'Ditch the algorithm' is the future of political protest. *The Guardian*. Retrieved from https://www.theguardian.com/commentisfree/2020/aug/19/ditch-the-algorithm-generation-students-a-levels-politics

Archer, H., Writer-Davies, R., & McGeoghegan, M. (2018). AI, automation, and corporate reputation. Retrieved from https://www.ipsos.com/en/ai-automation-and-corporate-reputation.

Ball, S.J., Junemann, C., & Santori, D. (2017). *Edu.net: Globalization and education policy mobility*. Routledge.

Bayne, S. (2015). Teacherbot: Interventions in automated teaching. *Teaching in Higher Education, 20*(4), 455-467. 10.1080/13562517.2015.102078

Bedingfield, W. (2020, Aug). Everything that went wrong with the botched A-Levels algorithm. *Wired*. Retrieved from https://www.wired.co.uk/article/alevel-exam-algorithm

Beer, D. (2018). *The data gaze: Capitalism, power and perception*. SAGE.

Bristows. (2018). Artificial Intelligence: Public perception, attitude and trust. Retrieved from https://www.bristows. com/viewpoint/articles/artificial-intelligence-public-perception-attitude-and-trust/

Brown, M. (2020). Seeing students at scale: How faculty in large lecture courses act upon learning analytics dashboard data. *Teaching in Higher Education, (25)*4, 384-400. 10.1080/13562517.2019.1698540

Bucher, T. (2016). Neither black nor box: Ways of knowing algorithms. In S. Kubitschko & A. Kaun (Eds.), *Innovative method in media & comm research* (pp. 81-98). Palgrave.

Bunz, M. (2017). The deed for a dialogue with technology. In M. T. Schafer & K. van Es (Eds.), *The datafied society: Studying culture through data* (pp. 249-254). Amsterdam University Press.

Chiusi, F., Fischer, S., Kayser-Bril, N., & Speilkamp, M. (Eds.). (2020). *Automating Society Report 2020*. Retrieved from https://automatingsociety.algorithmwatch.org

Christin, A. (2017). The mistrials of algorithmic sentencing. Logic, Issue 3 Justice. Retrieved from *https://logicmag.io/03-the-mistrials-of-algorithmic-sentencing/*

Council of Europe (2108). *Justice by algorithm – the role of artificial intelligence in policing and criminal justice systems. Motion for a recommendation* (Doc. 14628). Retrieved from http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=25062&lang=en

Dietvorst, B., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114-126. 10.1037/xge0000033

Dietvorst, B., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155-1170.

Edwards, R., & Fenwick, T. (2016). Digital analytics in professional work and learning. *Studies in Continuing Education, 38*(2), 213-227. 10.1080/0158037X.2015.1074894

Eicher, B., Polepeddi, L., & Goel, A. (2018). Jill Watson doesn't care if you're pregnant: Grounding AI ethics in empirical studies. *Proceedings of the In 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 88-94). ACM. Retrieved from https://doi.org/10.1145/3278721.3278760

Escobar, O. & Elstub, S. (2016). *Forms of mini-publics*. Research and Development Note 4. Retrieved from https://www.newdemocracy.com.au/2017/05/08/forms-of-mini-publics/

European Commission. (2019). *Ethics guidelines for trustworthy AI. Report by the High-Level Expert Group on Artificial Intelligence*. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Fenwick, T., & Nerland, M. (2014). Sociomaterial professional knowing, work arrangements and responsibility. New times, new cocepts? In T. Fenwick & M. Nerland (Eds.),

Reconceptualising professional learning: Sociomaterial knowledges, practices and responsibilities (pp. 1-7). Routledge.
Frontier Economics. (2018). *The Impact of Artificial Intelligence on Work*. Retrieved from https://royalsociety.org/topics-policy/projects/ai-and-work/
Gunning, D., & Aha, D. (2019). DARPA's Explainable Artificial Intelligence (XAI) program. *AI Magazine*, *40*(2), 44-58. https://doi.org/10.1609/aimag.v40i2.2850
Hao, K. (2019, January 21). AI is sending people to jail—and getting it wrong. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/
Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. SAGE.
Kolbjørnsrud, V., Amico, R., & Thomas, R. J. (2016). How artificial intelligence will redefine management. *Harvard Business Review, Nov*, 1-6.
Lange, A. C., Lenglet, M., & Seyfert, R. (2019). On studying algorithms ethnographically: Making sense of objects of ignorance. *Organization*, *26*(4), 598–617.
Lemay, D. J., Basnet, R. B., & Doleck, T. (2020). Fearing the robot apocalypse: Correlates of AI anxiety. *International Journal of Learning Analytics and Artificial Intelligence for Education, 2*(2), 24-33. 10.3991/ijai.v2i2.16759
Le Cun, Y, Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature, 521*, p. 436-444.
Luckin, R. & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach*. British Journal of Educational Technology, 50*(6), 2824-2838. 10.1111/bjet.12861
Ludvigsen, S. & Nerland, M. (2018). Learning at work: Social practices and units of analysis. In F. Fischer et al., (Eds.), *International Handbook of the Learning Sciences* (pp. 147-156). Routledge.
Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., Ko, R., & Sanghvi, S. (2017). *Jobs lost, jobs gained: Workforce transitions in a time of automation*. Retrieved from McKinsey & Company / McKinsey Global Institute website https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages /
Mateescu, A., & Nguyen, A. (2019). *Explainer: Algorithmic management in the Workplace*. Retrieved from Data & Society website http://datasociety.net
Networked Learning Editorial Collective (NLEC) (2020). Networked learning: Inviting redefinition. *Postdigital Science and Education*. Advance online publication. 10.1007/s42438-020-00167-8
Nilsson, N. J. (2010). *The quest for artificial intelligence: A history of ideas and achievements*. Cambridge University Press.
Ofqual. (2020). *Awarding GCSE, AS, A level, advanced extension awards and extended project qualifications in summer 2020: Interim report* (Ofqual/20/6656/1). Retrieved from https://www.gov.uk/government/publications/awarding-gcse-as-a-levels-in-summer-2020-interim-report
Partnership on AI. (PAI, 2018). *AI, Labor, and Economy Case Studies*. Retrieved from PAI website https://www.partnershiponai.org/compendium-synthesis/
Polonski, V. (2019, January 9). People don't trust AI – here's how we can change that. *The Conversation*. Retrieved from https://theconversation.com/people-dont-trust-ai-heres-how-we-can-change-that-87129
Priestley, M., Shapira, M., Priestley, A., Richie, M., & Barnett, C. (2020). *Rapid review of national qualifications experience 2020. Final report*. Retrieved from Scottish Government website: https://www.gov.scot/publications/rapid-review-national-qualifications-experience-2020/

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *Why should I trust you? Explaining the predictions of any classifier*. Retrieved from: arXiv:1602.04938v1

Robb, D. A., Padilla, S., Methven, T. S., Liang, U. Le Bras, P., Howden, T., Gharavi, A., Chantler, M. J., & Chalkiadakis, I. (2018). *Issues affecting user confidence in explanation systems*. ReaLX 2018: The SICSA Reasoning, Learning and Explainability Workshop, Aberdeen, Scotland. Retrieved from http://ceur-ws.org/Vol-2151/Paper_S7.pdf

Rosenblat, A., & Stark, L. (2015). Uber's drivers: Information asymmetries and control in dynamic work. *SSRN Electronic Journal*. 10.2139/ssrn.2686227

The Royal Society. (2018). *AI narratives: Portrayals and perceptions of AI and why they matter*. Retrieved from The Royal Society website: https://royalsociety.org/topics-policy/projects/ai-narratives/

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education, 151*. Advance online publication. 10.1016/j.compedu.2020.103862.

Susskind, R., & Susskind, D. (2015). *The future of the professions: How technology will transform the work of human experts*. Oxford University Press.

Thiele, K. (2014). Ethos of diffraction: New paradigms for a (post)humanist ethics. *Parallax, 20*(3), 202–216.

Tromans, R. (2019, July 1). *Now French lawyers demand statistical data ban, following judges' move*. Retrieved from Artificial Lawyer website https://www.artificiallawyer.com/2019/07/01/now-french-lawyers-demand-statistical-data-ban-following-judges-move/

Venturini, T. (2010). Diving in magma: How to explore controversies with actor-network theory. *Public Understanding of Science, 19*(3), 258–273.

Whittaker, M. et al. (2018). *AI Now Report 2018*. Retrieved from AI Now website http://ainowinstitute.org/AI_Now_2018_Report.pdf

Yu, K., Berkovsky, S., Conway, D., Taib, R., Zhou, J., & F. Chen. (2018) Do I trust a machine? Differences in user trust based on system performance. In J. Zhou & F. Chen (Eds.), *Human and Machine Learning* (pp. 245-264). 10.1007/978-3-319-90403-0_12