Critical Digital Humanities and Machine Learning

Caroline Bassett
c.bassett@sussex.ac.uk
University of Sussex, United Kingdom

David M. Berry
d.m.berry@sussex.ac.uk
University of Sussex, United Kingdom

M. Beatrice Fazi
b.fazi@sussex.ac.uk
University of Sussex, United Kingdom

Jack Pay
jp242@sussex.ac.uk
University of Sussex, United Kingdom

Ben Roberts
b.l.roberts@sussex.ac.uk
University of Sussex, United Kingdom

Panel Statement

This panel undertakes a speculative and theoretical discussion of possible future directions for digital humanities work driven by what we call informating, augmenting and automating technologies in the digital humanities. The panel particularly examines the emergence of a new paradigm of artificial intelligence around machine learning, statistical techniques and textual interfaces; a paradigm that challenges the way in which we understand the provision of digital humanities technologies and infrastructures. We explore the debates over informating, augmenting and automating processes that are now starting to emerge in digital humanities, and the historical trajectory that led to the current rapid changes from computational techniques. By looking at how machine-learning infrastructures effect knowledge formations, we engage with these new knowledges and practices, and argue that digital humanities must seek to contest and transform particular institutional structures that are problematic for humanities scholarship.

Although differences have emerged within the digital humanities between "those who use new digital tools to aid relatively traditional scholarly projects and those who believe that DH is most powerful as a disruptive political force that has the potential to reshape fundamental aspects of academic practice" (Gold, 2012: x), it is still the case that, as a growing and developing disciplinary area, digital humanities has much opportunity for these disparate elements to work together. Not unlike differences between empirical and critical sociology in a previous iteration of a contestation over knowledge, epistemology, disciplinary identity and research, digital humanities as a discipline will be richer and more vibrant with alternative voices contributing to projects, publications and practices. Indeed, the debates within digital humanities "bear the mark of a field in the midst of growing pains as its adherents expand from a small circle of like-minded scholars to a more heterogeneous set of practitioners who sometimes ask more disruptive questions" (Gold, 2012: x-xi).

Developing a critical approach to machine learning, for example, calls for computation itself to be historicised, and its developing relationship with humanities to be carefully uncovered. Similarly, by focusing on the materiality of machine learning, our attention is drawn to the microanalysis required at the level of computational conditions of possibility, combined with a macroanalysis of deployment of machine-learning systems in humanities work. This calls for us to think critically about how machine learning is being designed and deployed in the specific problem domains represented by the informating, augmenting and automating of digital humanities. The panel critically engages with these three modes of thought and practice, in order to connect and explore the present and possible future of digital humanities. We develop this approach in the context of these new techniques of knowledge-presentation, new infrastructures for knowledge work, and new formations around human capacities to work with complex and large data sets. Strong claims are often made about the potential for replacing aspects of traditionally humanities work undertaken by human labour alone through machine-learning techniques. Here, through a critical examination of new epistemologies and machine-generated data ontologies, for example, we examine the possibility of methods for a *critical digital humanities* in relation to new machine-learning techniques, together with how machine learning might be repurposed *for* a critical project within DH scholarship.

Towards a critique of machine learning: critical digital humanities and AI

David M. Berry

In this paper I investigate the claims of computational models and practices drawn from the field of artificial intelligence and more particularly machine learning. I do this to explore the extent to which machine learning raises important questions for our notions of being human, but also, relatedly the concept of civil society and democracy as distilled through notions of hermeneutic practice. That is, that in the 21st century we are seeing the creation of specific formations which threaten historical notions of humanities research and thinking. They represent new modes of knowing and thinking driven by these new forms of computation such as machine learning and Big Data, and which will have implications for the capacity to develop and use social and human faculties.

It is certainly the case that through the innovative assembling and organisation of scale technologies together with human actors new cognitive forms are under construction and experimentation. This paper develops a speculative and theoretical discussion of possible future directions driven by what we call Informating or Augmenting technologies in the digital humanities. In this paper, the notion of a digital humanities is linked to the social, cultural, economic and political questions of a recontextualisation and social re-embedding of digital technologies within a social field. Indeed, exploring the digital humanities through a critical lens I seek to understand how different disciplinary specialisms are newly refracted not just by their interaction, but also by the common denominator and limitations of computation. That is, how the constellation of concepts that are used within a disciplinary context are challenged and transformed within a computational frame.

Indeed, this raises theoretical and methodological questions, for example, digital humanities is keen develop tools to explore the new techniques such as machine learning for the field. This calls for a critical response, and there has already been some valuable work undertaken in this area, such as Alan Liu's work on critical infrastructure studies, but here I explore how a critical digital humanities can offer a way of thinking about the theoretical and empirical approach to massive-scale technologies. In this paper I argue that digital humanities should not only map these challenges but also propose new ways of reconfiguring research and teaching to safeguard critical and rational thought in a digital age. First, I turn my attention to research infrastructure and how critical approaches can contribute to and offer methods for contesting ML. I

argue that research infrastructures provide the technical a priori for the support of and conditions of possibility for digital humanities projects, but in a machine-learning paradigm different techniques and critical methods will be required to make sense of their use. Secondly, in relation to data, we might consider the more general implications of datafication not just within the general problem of big-data, but in terms of the specific issues raised by machine learning in the generation, processing and automated classification of data–especially where the metadata becomes nonhuman-readable.

This links to my final question about how visibility is made problematic when mediated through computational systems. The question is also linked to *who* and *what* is made visible in these kinds of machine-learning systems, especially where as Feminist theorists have shown, visibility itself can be a gendered concept and practice, as demonstrated in the historical invisibility of women in the public sphere, for example (see Benhabib, 1992). Finally, this paper will explore how to embed the capacity for reflection and thought into a critically-oriented digital humanities and thus to move to a new mode of experience, a two dimensional experience responsive to the potentialities of people and things intensified by the advances in machine-learning capacities. In other words, the reconfiguring of quantification practices and instrumental processes away from domination (Adorno, Horkheimer, Marcuse) and control (Habermas), instead towards reflexivity, critique and democratic practices. As Galloway argues, "as humanist scholars in the liberal arts, are we outgunned and outclassed by capital? Indeed we are–now more than ever. Yet as humanists we have access to something more important…. continue to pursue the very questions that technoscience has always bungled, beholden as it is to specific ideological and industrial mandates" (Galloway, 2014: 128). I argue that specific intervention points within the materialisation of this ML a priori, such as in design processes, can be explored to contest machine-learning techniques that serve to instrumentalise humanities approaches.

Digital humanities has the technical skills and cultural capital to make a real difference in how these machine-learning projects are developed, the ways in which instrumental logics are embedded within them and interventions made possible. For example, digital humanities through its already strong advocacy of open access, could push for and defend open source, open standards and copyleft licenses for technical components and software, opening up and document-

ing new techniques for machine-learning by humanists for humanists–but this could also be the opening up of the complexity of the black box of ML systems. The ways in which these aspects interrelate in terms of the ML "space of work" is hugely important, that is, the functional capacity of a machine-learning system is crucial, in as much as the range of humanities work may be adversely effected or inhibited by the shape of a machine-learning infrastructural system. I argue that these are urgent questions, with the recent turn towards what has come to be called "platformisation", that is the construction of a single digital system that acts as a technical monopoly within a particular sector, and it is certainly the case that the implications of machine-learning infrastructures and their black-boxed techniques for sorting, classification and ordering large amounts of data needs constant vigilance from digital humanists.

Augmenting and automating human and machine attention in the (digital) humanities

M. Beatrice Fazi

Attention denotes the cognitive process of selecting and focusing upon certain aspects of information whilst ignoring others. In recent years, it has been argued that this special state of percipient awareness is undergoing a profound transformation, due to the increasing intertwining of digital devices and everyday cognitive tasks (see Carr, 2010; Gazzaley and Rosen, 2016). Social media, phone apps, design interfaces, smart devices: the industry markets these technologies as helpful assistants that will free us from the chore of identifying, selecting and retaining relevant information, thus allowing us to dedicate our time, and our mental efforts, to other things. In addition, digital software and hardware are equally used to tune senses and to maintain motivation. Whilst cognitive cognates such as memory and intelligence are of course also targeted, it is the capacity to pay attention that seems primarily to be called into question here. In an attention economy, attention is believed to be a scarce commodity. The assumption is that, with the current information overload, digital machines are instruments able to outsource decisions regarding what to prioritise, what to select and what to discard in the data-deluge. Whilst much concern in the 20th century focused on the question "Can a machine think?", and Artificial Intelligence labs were devoted to answering this question, in the 21st century the central question seems to be "How do we think with machines, and how do we get machines to do much of our thinking for us?"

The exteriorisation of cognitive faculties such as the capacity for attention does, however, come at a price. Studies in neuroscience and neuropsychology, drawing from theories of neuronal plasticity, show that our brain is being rewired in favour of new cognitive skills, and to the detriment of older but cherished abilities, such as the capacity to read a novel from cover to cover. Evidence of this deterioration of human attention comes from science, yet everyday anecdotal confirmations also come from educators and parents, who report of children who cannot focus, and of students who are distracted and cannot complete their assignments. Relevantly, N. Katherine Hayles (2007) has described this situation in terms of a generational cognitive shift.

The humanities, due to the fact that they are largely based around texts, have often elaborated and developed concerns about human attention under the rubric of debates as to what counts as reading. Within the digital humanities, more specifically, it has been stressed that, whilst humans are very good at "close reading" (i.e. the careful, attentive and sustained inspection of a text), computing machines allow us to consider a broader picture. Franco Moretti (2013), amongst others, has called this condition "distant reading". These debates have opened up considerations about the possibility of an "algorithmic criticism" (Ramsay, 2011), as well as reflections on the importance of the hermeneutic faculties of human beings (Berry, 2012; Stiegler, 2010 and 2016). In this paper, I depart from these discussions within the digital humanities and then move to argue how new understandings of human attention might emerge in conjunction with possible conceptions of what I call "machine attention". I will map these possible conceptions of machine attention in relation to increasingly popular Artificial Intelligence techniques known as machine learning. More specifically, I will consider how machine-learning programs might be said or seen to pay attention to data-stimuli: they detect some information and discard some others, forming and dissolving patterns, in order to shape and sharpen their cognitive outcomes based on these selections. I will then emphasise the relevance of these modes of machine attention for the way in which we can understand what human attention might become after the computational turn in the humanities.

The question of what is happening to human attention is an important and pressing one for the humanities. It is always difficult to define what the humanities are, or where they begin and end. However, surely few would object that the humanities are the locus of

"deep" attention: humanities disciplines prioritise textual analysis, where the process of knowing is intimately connected to those of making sense, interpreting, and of giving meaning. These are epistemic processes that start and end with the cognitive exercise of attention. The pedagogical issue of what happens to students if they have lost (or never gained) the capacity to focus is also a question upon which the future of humanities disciplines, and humanities departments, seems to be predicated. In this paper I will address these issues, by considering the intermeshing of human and machine modes of attention, whilst also arguing that our engagement with automated forms of attention (as well as of other automated and augmented cognitive processes) should involve a commitment to re-defining and enlarging the prospect of what computational mechanisms are, and what rule-based, computational cognitive processes might amount to.

'The new spirit of automation': the changing discourse of automation anxiety

Caroline Bassett, Ben Roberts and Jack Pay

From self-driving cars, through high-frequency trading to military drones and organised swarms of shelf-stacking robots, our era is marked by rising automation and a new fascination with the likely social, cultural, and economic impacts of this computationally driven transformation. This paper will explore innovative methods by which the humanities might address contemporary *automation anxiety*. The wider topic of automation is a pressing subject with various existing academic responses such as Frey and Osborne's work on automation and the future of employment (2013). The focus of this paper is to address, as a topic in its own right, the cultural and social anxiety generated by these new forms of computational automation. What new research methods can the humanities use to map and understand automation anxiety around opaque computational decision making? What digital tools can be brought to bear on the diverse types of online public culture in which this anxiety is expressed?

Automation anxiety is evident in a plethora of popular contemporary accounts, public debates and political interventions. Tyler Cowen's *Average is Over* depicts a dystopian future in which the job market is divided between a highly educated and skilled elite capable of harnessing automation for personal wealth creation and a wider mass who are consigned to low paid work. Other accounts see in this new wave of computerisation the potential for a productive re-definition of the relationship with work. Futurists

Martin Ford in *Rise of the Robots* (2015) and Jerry Kaplan in *Humans Need Not Apply* (2015) propose to respond to the automation of work through the creation of a universal income. In a more radical version of this thesis, postcapitalism, as charted by Paul Mason, posits automation as the basis of a technologically-driven, non-market successor to capitalism. Another type of anxiety arises out of the increasing use of computerisation in law enforcement and military action. Here there is an automation anxiety that the current wave of military drones will evolve into fully autonomous killing machines, with software systems governing decisions about life and death. In July 2015 over 3000 robotics and artificial intelligence researchers and over 17000 other academics and interested parties (including Stephen Hawking, Elon Musk and Noam Chomsky) signed an open letter, published on the Future of Life website and widely disseminated in the global media, calling for a global ban on "offensive autonomous weapons beyond meaningful human control." There is also a more general anxiety which asks what happens to human life when so many tasks are automated away. Nicholas Carr's *The Glass Cage: Where Automation is Taking Us* (2014) suggests that automation is a threat to humanity itself—as we delegate tasks to computational tools, human cognitive capacities atrophy, understanding weakens, and the power of human reasoning is undermined.

This paper places contemporary automation anxiety in the context of historical debates about automation. It examines methods that might be used to analyse changing social attitudes to automation and computation between the 1960s and the present. Automation was a controversial topic in both Britain and the United States in the 1960s. In 1964 defence automation specialist Sir Leon Bagrit gave the public BBC Reith lectures on the topic. In the same year, President Lyndon B. Johnson set up the National Commission on Technology, Automation, and Economic Progress. Then, as now, there were concerns about automation and the future of employment. Then, as now, there were utopian imaginings of the future social benefits of automation. Nevertheless the hypothesis here would be that there are important differences between the two eras and that we can learn from changing attitudes to automation and computation. Among other things, analysis of changing attitudes to automation might illuminate different historical perspectives on: the end(s) of work; the relationship between labour and the domestic sphere; the role of computation in society.

The paper takes inspiration, but not theoretical orientation, from Boltanski and Chiapello's *The New Spirit of Capitalism* which used textual analysis of management literature from the 1960s and 1990s to argue for a fundamental shift in what they call the 'spirit of capitalism', i.e. the way in which capitalism justifies itself. In a similar vein we analysed key grey literatures (policy, commercial reports, academic and government papers) on automation from the 1960s and present in order to understand the changing discourse around automation. A key concern was to use digital humanities tools which provide different scales of analysis and new perspectives. We did this both to generate new understandings of automation anxiety across time and to investigate ways in which digital humanities and media archaeological approaches intersect.

The "new spirit" of capitalism which has emerged between the 1960s and the present day consists in a highly decentralised networked form of capitalism, characterised by "flatter" organisational hierarchy, much greater autonomy within firms for both individuals and teams, lower job security and the proliferation of temporary contracts and outsourcing. Boltanski and Chiapello use their sociological analysis of management literature to support a more speculative, philosophical account of capitalism and its critique, notably seeing the contemporary spirit of capitalism as incorporating the critiques that were made of capitalism in the 1960s and particularly around May 1968.

Similarly this paper argues that automation controversies could be a springboard to more general debates about the changing relationship between computation and society. The central premise here is that there is a much to be discovered from *attitudes to* automation and the justification of computation tools as there is from the specific technological forms and implementations of automation.

Bibliography

Bagrit, L. (1965). *The Age of Automation*. The BBC Reith Lectures 1964. London: Weidenfeld and Nicholson.

Benhabib, S. (1992). *Situating the Self: Gender, Community and Postmodernism in Contemporary Ethics*. London: Routledge.

Berry, D. M. (2011). *The Philosophy of Software: Code and Mediation in the Digital Age*. London: Palgrave Macmillan.

Berry, D. M. (2012). *Understanding Digital Humanities*. Basingstoke: Palgrave.

Berry, D. M. (2014). *Critical Theory and the Digital*. New York: Bloomsbury

Berry, D. M. and Fagerjord, A. (2017). *Digital Humanities: Knowledge and Critique in a Digital Age*. Cambridge: Polity.

Boltanski, L. and Chiapello, E. (2007). *The New Spirit of Capitalism*. London: Verso.

Carr, N. (2010). *The Shallows: What the Internet is Doing to Our Brains*. New York: W. W. Norton.

Carr, N. G. (2014). *The Glass Cage: Automation and Us*. New York: Norton.

Cowen, T. (2013). *Average Is Over: Powering America Beyond the Age of the Great Stagnation*. New York: Dutton.

Ford, M. (2015). *The Rise of the Robots: Technology and the Threat of Mass Unemployment*. London: Oneworld Publications.

Frey, C. B. and Osborne, M. A. (2013). *The Future of Employment: How Susceptible Are Jobs to Computerisation*. Oxford Martin Working Papers.

Kaplan, J. (2015). *Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence*. New Haven, CT: Yale University Press.

Galloway, A. (2014). "The cybernetic hypothesis." *differences*, 25(1): 107-31.

Gazzaley, A. and Rosen, L. D. (2016). *The Distracted Mind: Ancient Brains in a High-Tech World*. Cambridge, MA: The MIT Press.

Gold, M. K. (2012). "The digital humanities moment." In M. K. Gold (ed) *Debates in the Digital Humanities*. Minneapolis, MI: University of Minnesota Press, pp. ix-xvi.

Hayles, N. K. (2007). "Hyper and deep attention: the generational divide in cognitive modes." *Profession 2007*, pp. 187-99.

Mason, P. (2015). *Postcapitalism: A Guide to Our Future*. London: Allen Lane.

Moretti, F. (2013). *Distant Reading*. London: Verso.

Ramsay, S. (2011). *Reading Machines*: *Toward an Algorithmic Criticism*. Champaign, IL: University of Illinois Press.

Stiegler, B. (2010). *Taking Care of Youth and the Generations*. Translated by S. Barker. Stanford, CA: Stanford University Press.

Stiegler, B. (2016). *Automatic Society*. *Volume 1: The Future of Work*. Cambridge: Polity.