

RESEARCH ARTICLE

Bell Labs and the ‘neural’ network, 1986–1996

Harry Law

Department of History and Philosophy of Science, University of Cambridge, UK
Corresponding author: Harry Law, Email: hp464@cam.ac.uk

Abstract

Between 1986 and 1996 researchers at the AT&T Bell Laboratories Adaptive Systems Research Department curated thousands of images of handwritten digits assembled by the United States Postal Service to train and evaluate artificial neural networks. In academic papers and conference literature, and in conversations with the press, Bell Labs researchers, executives and company spokespeople deployed the language of neurophysiology to position the systems as capable of codifying and reproducing feats of perception. Interpretations such as these were pivotal to the formation of brain–computer imaginaries that surrounded the development of the systems, which obscured the institutional infrastructures, clerical and cognitive labour, and the manipulation and maintenance of data on which feats of ‘recognition’ depended. Central to building the group’s networks was the development of data sets constructed in consort with the US Postal Service, which arbitrated between the practicality of conducting research and the representation of an extensive catalogue of possible forms and permutations of handwritten digits. These imaginaries, which stressed a likeness with the human brain, were compounded by the promotion of ‘successful applications’ that took place under the AT&T corporate umbrella and with the essential support of US Postal Service workers to correct system errors.

This paper considers efforts by AT&T Bell Laboratories to develop, evaluate and deploy artificial intelligence (AI) systems capable of optical character recognition (OCR), a process by which software recognizes letter shapes in an image to generate a corresponding text file.¹ To perform this procedure, Bell Labs computer scientists utilized artificial neural networks, a form of statistical learning linked to the connectionist school of AI developed in the early to mid-twentieth century in which interconnected processing nodes are used to generate a predictive function by analysing underlying relationships in a set of data.² Trained on, and assessed by, their ability to correctly identify digits within up to 10,000 images of public letters collected, sorted and supplied by the US Postal Service (USPS), the systems determined ‘features’ (in this context, a piece of information about the content of an image) such as closed loops, line direction and line intersections to identify handwritten digits from within the range of zero to nine. While the group later partnered with the National Institute of Standards and Technology to develop

1 Mara Mills, ‘Beyond recognition: what machines don’t read’, *American Foundation for the Blind Blog*, at www.afb.org/blog/entry/beyond-recognition-what-machines-dont-read (accessed 20 February 2021).

2 J. Mark Bishop, ‘History and philosophy of neural networks’, in Hisao Ishibuchi (ed.) *Computational Intelligence*, Paris: Eolss Publishers, 2015 pp. 22–9.

data sets for building its systems, this paper is focused solely on networks designed and tested using data supplied by the USPS. My analysis foregrounds the role of brain-computer imaginaries in the development and popularization of systems built for these purposes by considering the ‘computational metaphor’ in neuroscience, which involves drawing parallels between the functioning of computers and the brain.³ The computer scientist Edsger W. Dijkstra has argued, ‘A more serious byproduct of the tendency to talk about machines in anthropomorphic terms is the companion phenomenon of talking about people in mechanistic terminology.’⁴ The result, as described by Alexis T. Baria and Keith Cross, is that ‘the human mind is afforded less complexity than is owed, and the computer is afforded more wisdom than is due.’⁵ This paper follows in that tradition to argue that brain-computer imaginaries concealed the nature of the group’s artificial neural networks as complexes of technical and clerical labour, institutional proficiency, and the careful collection, processing and curation of data.

My analysis begins by considering Bell Labs’ position within the AT&T group following a landmark 1982 divestiture ruling before introducing the Adaptive Systems Research Department and briefly situating the group’s work within the broader genealogy of AI research. Properly contextualized, the paper turns to the group’s early forays into artificial neural-network research between 1986 and 1990 to propose that computer vision systems were introduced to confirm the promise of connectionism by evoking metaphors of cognition and of the brain. Next, I consider 1990–2, a period in which research activity intensified as the group developed sophisticated hardware and software with the goal of improving performance. I connect the formation of brain-computer imaginaries with the obfuscation of clerical and cognitive labour provided by researchers, contractors and US Postal Service workers. In my third and final section, I demonstrate that carefully chosen commercial applications naturalized the institutional infrastructures from which the predictive power of the artificial neural network was derived. In doing so, I illustrate that the manner of the group’s research and the commercial deployments that it enabled concealed an extensive web of contingencies, facilitating – through the formation of imaginaries that stressed a likeness with the human brain – their portrayal as autonomous entities capable of perception and recognition.

To make this case, I draw on the ‘proxy’ as described by historian Dylan Mulvin in *Proxies: The Cultural Work of Standing In*. Mulvin argues that the ‘proxy’ functions as the ‘necessary form of make-believe and surrogacy that enable[s] the production of knowledge’.⁶ Proxies exist as ‘working objects’ as defined by Peter Galison and Lorraine Daston, which act as simplified representations of complex systems, structures and phenomena.⁷ Historian Nathan Ensmenger, for example, argues that the choice of chess as an experimental medium by AI researchers in the mid- to late twentieth century redefined how chess was played, and what it meant to be an ‘intelligent’ chess player.⁸ My analysis is focused on the group’s artificial neural networks and their application to computer vision tasks. Such networks, as I propose in what follows, existed as assemblies of institutional proficiency; technical, menial and clerical labour; and the gathering, refining and

3 R.M. Biron, ‘The computational metaphor and computer criticism’, *Journal of Computing in Higher Education* (1993) 5(1), pp. 111–31.

4 E. Dijkstra, ‘On anthropomorphism in science. Philosophers’ Lunch, Austin, TX’, (25 September 1985), at www.cs.utexas.edu/users/EWD/ewd09xx/EWD936.PDF.

5 Alexis T. Baria and Keith Cross, ‘The brain is a computer is a brain: neuroscience’s internal debate and the social significance of the Computational Metaphor’, *ArXiv*, 2021, pp. 1–14, 4.

6 Dylan Mulvin, *Proxies: The Cultural Work of Standing In*, Cambridge, MA: MIT Press, 2021, pp. 4–5.

7 Lorraine Daston and Peter Galison, *Objectivity*, New York: Zone Books, 2007, pp. 21–2.

8 Nathan Ensmenger, ‘Is chess the drosophila of artificial intelligence? A social history of an algorithm’, *Social Studies of Science* (2012) 42(1), pp. 5–30.

treatment of data for development and assessment purposes. I argue that the sociotechnical complexion of these systems was concealed by both the discourse of neuroscience and, with respect to the data on which they were trained and tested, a related ability to moderate between the practicality of conducting research and the representation of an extensive catalogue of possible forms and permutations of handwritten digits.

My account, which links technical practice to the erasure of labour, mirrors interventions in the history of computing that have considered why and how labour, origins and contingencies become hidden. Mar Hicks has studied the development of the computing industry in the United Kingdom in the mid- to late twentieth century by drawing on theories of structural sexism to explain the erasure of women's labour.⁹ Neda Atanasoski and Kalindi Vora have formulated a theory of labour erasure grounded in the dehumanizing practices of colonialism by extending analyses, in critical ethnic studies, of gendered racialization to include machine 'others'.¹⁰ More recently, Lilly Irani has studied the Amazon Mechanical Turk platform to argue that contemporary accounts of human computation rely on worker invisibility, while Kate Crawford has rendered visible the role of low-wage information workers in enabling the functioning of AI systems.¹¹ Accounts by Lucy Suchman and David Mindell propose that the boundary between machine and human is neither separate, nor impermeable, nor stable, and that narratives and frameworks of 'autonomy' are easily deployed to erase the imprint of labour.¹² By investigating the discourse of neuroscience and a connected capacity for artificial neural networks to act as stand-ins for fragments of the world, my analysis seeks to contribute to this body of scholarship by foregrounding the role of labour amongst other overlooked contingencies in the development and deployment of Bell Labs' computer vision systems.

In the years prior to the formation of the Adaptive Systems Research Department, the AT&T Corporation signed a divestiture agreement with the federal government of the United States mandating the break-up of the company in January of 1982. Superseding a 1956 antitrust deal, the agreement compelled AT&T to relinquish control of the regional Bell Operating Companies that provided local telephone services in the United States and Canada.¹³ The terms of the deal stated that AT&T's regional firms, or so called 'Baby Bells', would no longer be supplied with equipment from AT&T subsidiary Western Electric, a move designed to promote competition in the supply of telephone equipment to service providers.¹⁴ In return AT&T retained control over subsidiaries such as Western Electric, and, in a major coup, successfully lobbied the federal government to quash the terms of a 1956 antitrust consent decree that barred the firm from participating in the general sale of computers. Free to translate computing research into products such as lightwave systems and fibre optics, Bell Telephone Laboratories reorganized as AT&T Bell Laboratories in 1984 with a focus on the technologies that the firm believed would define

⁹ Mar Hicks, *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*, London: MIT Press, 2017.

¹⁰ Neda Atanasoski and Kalindi Vora, *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*, Durham, NC: Duke University Press, 2019.

¹¹ Lilly C. Irani and M. Six Silberman, 'Turkopticon: interrupting worker invisibility in Amazon mechanical Turk', in *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), pp. 611–20; Kate Crawford, *Atlas of AI*, New Haven, CT: Yale University Press, 2021.

¹² Lucy Suchman, *Human-Machine Reconfigurations*, Cambridge: Cambridge University Press, 2021; David Mindell, *Our Robots, Ourselves: Robotics and the Myths of Autonomy*, New York: Viking, 2015.

¹³ Jeanne Saddler, 'Limit on 1984 rise in local phone bills clears house over opposition by AT&T', *Wall Street Journal*, 11 November 1983, p. 4.

¹⁴ Joseph D. Kearney, 'From the fall of the Bell system to the Telecommunications Act: regulation of telecommunications under Judge Greene', *Hastings Law Journal* (1999) 50(6), pp. 1395–1471.

the so-called ‘Information Age’ of the future.¹⁵ The technologies of that tomorrow were, however, tied firmly to the commercial realities of the day.

Reflecting on the reorganization of AT&T, Robert Lucky, a Bell Labs executive from 1983 to 1991, suggested that the divestiture created financial pressures that forced the firm to switch its focus away from ‘fundamental research’ and towards commercial applications. Lucky told the historian David Hochfelder in 1999, ‘With the passing of the old Bell Labs model the one percent license contract fee disappeared, and with it the blue sky thinking ... the singular event that cracked that open was the antitrust trial.’¹⁶ Underlining the new financial reality facing the company, the *Washington Post* estimated that AT&T’s assets totalled approximately \$152 billion in 1983, falling to \$34 billion in 1984.¹⁷ *Time* magazine similarly reported that the firm held \$155 billion in assets before the divestiture and \$35 billion after its completion.¹⁸ A commercial imperative to develop the computational technologies of the ‘future’ – and a licence from the federal government to do so – was therefore the backdrop against which Bell Labs embarked on its programme of research into artificial neural networks designed for computer vision applications.

Beyond recognition, 1986–1990

The Adaptive Systems Research Department based in Holmdel, New Jersey was founded in 1984 by Richard E. Howard, head of the Microelectronics Research Department, and Lawrence ‘Larry’ Jackel, head of the Device Structure Research Department, at Bell Labs.¹⁹ The goal of the unit was, according to AT&T, to ‘build large analogues of neural networks by creating highly connected silicon networks that process information continuously and collectively’.²⁰ Research of this type is often situated within the connectionist school of artificial intelligence, which popular histories of AI suggest was displaced by its symbolic counterpart (in which systems are developed using hard-coded rules based on the manipulation of expressions) in the ‘winter’ of the 1960s before returning to prominence in the ‘summer’ of the 1980s.²¹ Though contested by scholars who stress the need to consider the material, political and social circumstances surrounding computation and data, accounts of this resurgence in folk histories typically favour the role of technical advances such as backpropagation, which was linked with character recognition applications and pursued independently by computer scientists including Geoffrey Hinton and Terrence Sejnowski, David Rumelhart and – as we shall see – Bell Labs researcher Yann LeCun.²² Like many others working on connectionist systems during this period, the New Jersey group’s character recognition networks used data provided by the USPS split into a training set used to develop the network and a test set used to assess its performance.

Although formed in 1984, the first major piece of artificial neural-network research from the group was published in 1987 by hardware specialist Hans-Peter Graf and department leads Larry Jackel and Richard Howard. The period included four significant

15 AT&T promotional video, *The Information Age*, 1985, AT&T Archives and History Center, New Jersey.

16 David Hochfelder, ‘Robert Lucky: an interview conducted by David Hochfelder’, 10 September 1999, Interview #361 for the IEEE History Center, the Institute of Electrical and Electronics Engineers, Inc.

17 Caroline E. Mayer and Merrill Brown, ‘A new era of hot competition’, *Washington Post*, 12 December 1983.

18 Anon., ‘Click! Ma is ringing off’, *Time*, 21 November 1983, pp. 61–3.

19 Y. LeCun, L.D. Jackel, B. Boser, J.S. Denker, H.P. Graf, I. Guyon, D. Henderson, R.E. Howard and W. Hubbard, ‘Handwritten digit recognition: applications of neural net chips and automatic learning’, *IEEE Communications Magazine* (1989) 27(11) pp. 41–6.

20 AT&T promotional video, *People in Touch with the Future*, 1987, AT&T Archives and History Center, New Jersey.

21 Daniel Crevier, *AI: The Tumultuous History of the Search for Artificial Intelligence*, New York: Basic Books, 1994.

22 Nils Nilsson, *The Quest for Artificial Intelligence*, Cambridge: Cambridge University Press, 2009.

milestones in computer vision research beginning in 1988 with the group's efforts to analyse handwritten digits authored by American researcher John Denker. This research was followed by a paper written in 1989 by French computer scientist Yann LeCun, who, after joining the firm in 1988, applied a version of the backpropagation algorithm (the automatic adjustment of the 'weights' in a multilayer neural network that transforms input data into outputs) to a system similar in design to the 'Neocognitron' artificial neural-network pattern recognizer developed by computer scientist Kunihiko Fukushima in 1980.²³ LeCun's backpropagation paper was followed by his development of the 'optimal brain damage' technique, which involved removing large numbers of surplus 'weights' in order to improve performance. The last major research paper in this period, published in 1989, included the application of both the backpropagation algorithm and the 'optimal brain damage' technique within a single system. All four research efforts included data provided by the US Postal Service, which was the largest employer in the United States during the period of my study.²⁴ I consider each milestone to argue that the 'perception' of digits was dependent on the nature of the networks as fusions of data, labour, expertise and institutional power whose make-up was hidden by the discourse of neuroscience and a capability to represent simplified analogues of complex systems, structures and phenomena.

Holmdel's programme of research began with attempts to forecast the ineffectiveness of non-neural-network computers to perform activities such as recognition and appeals to brain-computer metaphors that hailed artificial neural networks as a solution. Denker, Jackel and Howard wrote in a 1987 technical report, 'Even extending [non-neural-network computer] technology to these extremes, though, there are still problems like speech recognition or computer vision that appear to be beyond our computational abilities.' The solution, according to the researchers, would be found in 'developing model computation systems with some of the features of natural computers'.²⁵ Underlining early links between imaginaries of mind and machine in New Jersey, Denker and colleagues recast the brain as an 'organic computer' in a paper submitted to the inaugural 1987 Conference on Neural Information Processing Systems (NIPS), an interdisciplinary gathering of neuroscientists and machine-learning researchers.²⁶ The language illustrates that Denker, who in 1989 criticized 'beltway bandits' who argued that artificial neural networks might act as a 'general problem solver', was willing to indulge in brain-computer metaphors while expressly rejecting the notions of universality invigorated by the comparisons.²⁷ Indeed, in a colloquial interpretation of the group's research, the *Financial Times* reported in 1987 that Bell Labs had developed a "'neural network", a silicon chip which mimics the way some brain cells retrieve information and solve problems', while *Dallas Morning News* announced in the same year that the firm had created 'a chip that implements neural computing, a new approach to developing computers capable of performing tasks like the human brain'.²⁸ Both in the problem posed and in the solution offered, the group embarked on its programme of research by indulging in the

23 Kunihiko Fukushima, 'Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position', *Biological Cybernetics* (1980) 36, pp. 193–202.

24 James Bovard, 'The slow death of the U.S. postal service', *Cato Institute Policy Analysis* (1988) pp. 1–12.

25 Richard Howard, Lawrence Jackel and Hans Graf, 'Electronic neural networks: a new class of computing hardware', in *ESSCIRC '87: 13th European Solid-State Circuits Conference* (1987), pp. 13–18.

26 Stuart Mackie, Hans P. Graf, Daniel B. Schwartz and John Denker, 'Microelectronic implementations of connectionist neural networks', in Dana Anderson (ed.), *Proceedings of the 1987 International Conference on Neural Information Processing Systems*, New York: American Institute of Physics, 1988, pp. 515–23.

27 Loring Wirbel, 'Neurals no panacea', *Electronic Engineering Times*, 30 January 1989, p. 54.

28 David Fishlock, 'Ma Bell's Christmas gift to mankind: the 40th birthday of transistors, the most pervasive invention since the wheel', *Financial Times*, 23 December 1987; Jim Mitchell, 'Working in a dream world: Bell Labs researchers labor in realm of the absurd', *Dallas Morning News*, 7 November 1987.

co-production of brain-computer imaginaries premised on the blending of the language of neurophysiology and computation.

With computer vision identified as a promising area for the application of artificial neural-network technology, Howard, Jackel, Denker and six other colleagues published in 1988 the results of a system designed to recognize the numbers between zero and nine by processing over 10,000 256-bit vectors consisting of raw bitmaps of ZIP codes on postage envelopes, which were collected, sorted and photographed by the USPS. The system achieved an ‘error rate’ (the proportion of incorrectly identified examples) of 6 per cent.²⁹ Though misclassifying 6 per cent of images – compared to a 2.5 per cent ‘error rate’ for the human eye estimated by the group – the 1988 research was deemed successful enough to warrant a comparison with the mammalian visual cortex.³⁰ Denker and his colleagues referenced a landmark 1962 study by neuroscientists David Hubel and Torsten Wiesel in which the pair mapped the functional architecture of the cat’s eye.³¹ Commenting on perceived similarities, Denker and colleagues wrote, ‘Some kernels [a type of filter used to extract features from an image] synthesized by the network can be interpreted as feature detectors remarkably similar to those found to exist in biological vision systems.’³² In his account of acoustic research at AT&T in the early twentieth century, Jonathan Sterne argued that notions of perceptual processes are coded in technological development by exploring the relationship between telephone research and the problems, materials and methods of hearing research.³³ This reflexivity was in play during the 1980s, with AT&T’s ‘electronic ear’ making way for the artificial eye. Electronic eyes, however, were only part of the story. A second 1988 paper detailing LeCun’s ‘optimal brain damage’ technique demonstrated that by removing certain ‘weights’ (values that control the strength of the connection between two units in a network), systems can typically expect a superior generalization ability.³⁴ Mara Mills has suggested that hearing impairment was used to formulate the compressed signals of the telephone system developed by AT&T in the 1940s.³⁵ Phenomena of this type, described by Mills and Sterne as ‘dismediation’, describe disability as a constituting dimension of technology, and technology as a constituting dimension of disability.³⁶ In this way, amidst the borrowed rhetoric of neuroscience, the characterization of ‘optimal’ brain damage signals a process by which conceptions of cognitive disability – and with it, neurophysiology – became entangled with technological practice.

In 1988, LeCun was the lead author on a paper that described the application of a version of the backpropagation algorithm to Denker’s ZIP code classification research.

29 J.S. Denker, W.R. Gardner, H.P. Graf, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, H.S. Baird and I. Guyon, ‘Neural network recognizer for hand-written zip code digits’, in David Touretzky (ed.), *Proceedings of the 1st International Conference on Neural Information Processing Systems*, Cambridge, MA: MIT Press, 1988, pp. 323–31, 330.

30 Private communication from E. Sackinger and J. Bromley, cited in Vladimir Vapnik, ‘Principles of risk minimization for learning theory’, in John E. Moody (ed.), *Proceedings of the 4th International Conference on Neural Information Processing Systems*, San Francisco: Morgan Kaufmann Publishers, 1991, pp. 831–8, 838.

31 David Hubel and Torsten Wiesel, ‘Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex’, *Journal of Physiology* (1962) 160(1), pp. 106–54.

32 Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard and L.D. Jackel, ‘Backpropagation applied to handwritten zip code recognition’, *Neural Computation* (1989) 1(4), pp. 541–51, 547.

33 Jonathan Sterne, *MP3: The Meaning of a Format*, Durham, NC: Duke University Press, 2012, p. 33.

34 Yann LeCun, John S. Denker and Sara A. Solla, ‘Optimal brain damage’, in David Touretzky (ed.), *Proceedings of the 2nd International Conference on Neural Information Processing Systems*, Cambridge, MA: MIT Press, 1988, pp. 598–605, 602.

35 Mara Mills, ‘Deafening: noise and the engineering of communication in the telephone system’, *Grey Room* (2011) 43, pp. 118–45, 135.

36 Mara Mills and Jonathan Sterne, ‘Afterword II: dismediation – three proposals, six tactics’, in Elizabeth Ellcessor (ed.), *Disability Media Studies*, New York: NYU Press, 2017, pp. 365–78.

Informally dubbed ‘LeNet’, the system was trained on images provided by the US Postal Service subject to a process known as ‘skeletonization’ in which unusual characteristics were removed from the images. Researchers noted that ‘extraneous marks’ were eliminated, while a scaling process known as ‘linear transformation’ was performed to ensure that all images were the same size.³⁷ A segmentation process, in which ZIP codes were separated into single digits, was conducted by Postal Service contractors and State University of New York researchers Ching-Huei Wang and Sargur N. Srihari. The contractors developed a computational framework for separating individual digits on envelopes from the ZIP code, which the group used to train artificial neural networks between 1988 and 1996.³⁸ LeCun and colleagues reported an ‘error rate’ of 5 per cent in their paper, while characterizing the exercises of segmentation, skeletonization and linear transformation as representative of ‘minimal preprocessing’.³⁹ Despite this comment, LeCun, Denker, Howard and Jackel later acknowledged, at least with regard to segmentation, that the process ought to be considered a ‘very difficult problem’.⁴⁰ In his history of Charles Babbage’s analytical engine, Simon Schaffer argues that fundamental to the persuasive capacity of the rhetoric of ‘intelligent’ machines is the obscuring of the role of labour.⁴¹ A similar relationship manifested in New Jersey during this period as the minimization of the processes performed by the US Postal Service contractors enabled LeNet’s acts of ‘recognition’ to be linked with the independent action of the artificial neural network. In this way, the contribution of technical and sub-technical labour was shaded by a reduction in the standardized ‘error rate’ used to demonstrate the effectiveness of the backpropagation algorithm.

From specialization to generalization, 1990–1992

Research activity intensified after 1990 as the group developed a flurry of artificial neural-network permeations based on, and assessed by, data provided by the US Postal Service to test and train the systems. Between 1990 and 1992 significant advances in the design of specialist hardware built to improve benchmark performance compounded early reductions in error rates. In what follows, I consider the way in which highly specific successes in analysing ‘atypical data’ were used to gesture towards evidence of reliability before considering early commercial applications and supporting comments made in the press in which Bell Labs spokespeople figuratively described industrial deployments as a means to ‘simplify’ work. The section concludes with an examination of Holmdel’s continued use of the language of neurobiology amidst a reduction in error rates associated with the US Postal Service data set used to train and assess the group’s artificial neural networks.

The success of the networks in isolated examples of challenging character recognition tasks was one way in which researchers signalled an ideal of synthetic ‘recognition’. LeCun, Denker, Jackel and six Bell Labs colleagues published in 1990 an updated version of LeNet, which was trained and tested on a database of 9,298 segmented numerals collected from ZIP codes that appeared on envelopes passing through the Buffalo, New York postal office. Despite changes to the design of the network, the researchers

³⁷ LeCun *et al.*, *op. cit.* (32), p. 542.

³⁸ Ching-Huei Wang and Sargur N. Srihari, ‘A framework for object recognition in a visually complex environment and its application to locating address blocks on mail pieces’, *International Journal of Computer Vision* (1988) 125(2), pp. 125–51.

³⁹ Yann LeCun *et al.*, *op. cit.* (32), p. 549.

⁴⁰ Yann LeCun *et al.*, *op. cit.* (32), p. 549.

⁴¹ Simon Schaffer, ‘Babbage’s intelligence: calculating engines and the factory system’, *Critical Inquiry* (1994) 2(1), pp. 203–27.

continued to use digits that had been collected and segmented by the US Postal Service before eliminating ‘extraneous marks’ and ‘normalizing’ each digit via the linear-transformation scaling process. Achieving a reduced error rate of 3.4 per cent, which LeCun and colleagues attributed to the network’s ‘*a priori* knowledge’ linked to the use of specialized feature detectors, the authors highlighted the system’s success in categorizing ‘atypical data’ in which handwritten digits were penned in a highly stylized manner as an indicator of dependable performance.⁴² Indeed, Jackel recorded a video of Yann LeCun and colleagues celebrating a network correctly identifying irregular examples of digits just two years after the development of the updated version of LeNet.⁴³ Descriptions of success in categorizing ‘atypical data’ used to infer reliable performance surface a vein of self-referential logic core to demonstrating the applicability and usability of Bell Labs’ computer vision systems. Data standing in for data, abstract edge cases were deemed to be representative of the systems’ brain-like capacity to simultaneously identify both unusual and model characteristics at the outset of the decade.

As error rates fell, the group turned its attention to commercial applications, which drew into focus the role of labour in maintaining the effective functioning of (and, accordingly, the imaginaries surrounding) its artificial neural networks. Automation was referenced in Holmdel’s research papers from the outset, with the group citing Princeton University computer scientist Theodosios Pavlidis and colleagues, who reflected that ‘variable size character recognition, especially for text that is mixed with graphics, would find many applications in office automation’.⁴⁴ Despite this, speaking to the *Los Angeles Times* in 1990, Bell Labs executive Robert Lucky described commercial deployments of character recognition systems in terms of simplification: ‘I think people want the world to be a simpler place ... we all yearn for simplification, but we’re all governed by the second law of thermodynamics.’⁴⁵ In an interview with *R & D Magazine* the following year, Bell Labs researcher Chris Burges euphemistically described the role of clerical labour in the correction of system errors as efforts to ‘incorporate contextual analysis’ to identify handwriting on postal envelopes.⁴⁶ Comments by Lucky and Burges belong to a well-established tradition of positioning forms of mechanization as a means to save labour, circumventing the notion that such moves tend to replace it – rather than supplement it. Lorraine Daston, in her history of mathematical calculation, has argued that formulations of this type are premised on the revaluation of certain sorts of cognitive labour and systems of recognition, and that the cognitive tasks associated with notions of intelligence reflect social and class hierarchies.⁴⁷ Indeed, with respect to clerical work in the twentieth century, Craig Robertson’s *The Filing Cabinet* has explored the way in which sorting was linked to technical practice, gender, labour and notions of information collection and circulation.⁴⁸ Interventions such as those made by Lucky and Burges, while mischaracterizing the essence of automation, acknowledged that in commercial environments the systems’ reputation for self-sufficiency was contingent on both the

42 Y. LeCun, O. Matan, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel and H.S. Baird, ‘Handwritten zip code recognition with multilayer networks’, in *10th International Conference on Pattern Recognition Proceedings* (1990) 2, pp. 35–40, 36.

43 Yann LeCun, ‘Convolutional network demo from 1993’, YouTube, at www.youtube.com/watch?v=FwFduRA_L6Q (accessed 20 February 2021).

44 S. Kahan, T. Pavlidis and H.S. Baird, ‘On the recognition of printed characters of any font and size’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1987, pp. 274–88, 274.

45 Michael Schrage, ‘Where technology is taking us’, *Los Angeles Times*, 7 January 1990.

46 Tim Studt, ‘Neural networks: computer toolbox for the ‘90s’, *R & D Magazine*, 1 September 1991.

47 Lorraine Daston, ‘Enlightenment calculations’, *Critical Inquiry* (1994) 21(1) pp. 182–202.

48 Craig Robertson, *The Filing Cabinet: A Vertical History of Information*, Minneapolis: University of Minnesota Press, 2021.

replacement of existing labour and the introduction of a new type of worker to ensure reliable operation.

Amidst a backdrop of early industrial deployments, between 1990 and 1992 the group focused efforts on the development of specialized hardware in order to boost performance, and, with it, the fabrication of imaginaries that magnified the biological connotations of the artificial neural network. In 1990 Graf and colleagues designed a bespoke chip for pattern recognition applications by analysing the performance of the group's artificial neural networks, explaining that they were 'computationally very expensive' and thus required the development of 'special-purpose hardware'.⁴⁹ A further two chips, a specialized design known as ANNA and an 'analogue neural-network processor', were developed in 1991.⁵⁰ In a 1991 interview with Bell Labs computer scientist Robert Frye, *R & D Magazine* journalist Tim Studt coarsely underlined the equivalence between the neural network and the brain: 'They're fast! They can read! They can generalize! They can do lots of different jobs. They think like you and me. They're neural networks.'⁵¹ By recasting computers as artificial brains, Holmdel's researchers presented the 'problem' of recognition to the public as one of scale. Appealing to perceived similarities between idealized electronic 'neurons' and their organic counterparts, the size of an artificial neural network – determined by limits in processing power and specialized hardware that Bell Labs was well placed to remove – was thus presented as the limiting factor in replicating a fuller breadth of human cognition.

'Proving grounds', 1992–1996

I turn now to the consolidation of reductions in error rates and an acceleration in industrial deployments of the group's systems by interrogating Holmdel's distinction between 'proving grounds' and 'successful applications'.⁵² In doing so, I demonstrate that internal deployments at AT&T and in partnership with companies owned by the telecommunications giant were used to prove the value of character recognition artificial neural networks throughout this period. The section concludes by illustrating that commercial applications continued to rely on the creation of new roles that corrected system errors before proposing that industrial deployments collided with increases in speed and performance to energize imaginaries of mind and machine in the middle of the decade.

With incremental gains on US Postal Service data sets secured, the group sought additional applications for its character recognition technology. Such applications, however, primarily consisted of deployments in conjunction with the US Postal Service, internally at AT&T or in partnership with companies owned by the firm. In a paper authored by Jackel and nineteen other researchers, the group wrote that 'Character Recognition has served as one of the principle proving grounds for neural-net methods and has emerged as one of the most successful applications of this technology'.⁵³ Indeed, 'proving grounds'

49 H.P. Graf, R. Janow, D. Henderson and R. Lee, 'Reconfigurable neural net chip with 32K Connections', in Moody, op. cit. (30), pp. 1032–8.

50 B.E. Boser, E. Sackinger, J. Bromley, Y. LeCun, R.E. Howard and L.D. Jackel, 'An analog neural network processor and its application to high-speed character recognition', *IJCNN-91-International Joint Conference on Neural Networks* (1991), pp. 415–20; E. Säckinger, B. Boser and L. Jackel, 'A neurocomputer board based on the ANNA neural network chip', in Moody, op. cit. (30), pp. 773–80, 780; Bernhard Boser, Eduard Säckinger, Jane M. Bromley, Yann Le Cun and Lawrence D. Jackel, 'An analog neural network processor with programmable topology', in *IEEE Journal of Solid-State Circuits* (1993) 26(12), pp. 2017–25.

51 Studt, op. cit. (46).

52 Lawrence Jackel et al., 'Neural-net applications in character recognition and document analysis', in Ben Yuhás and Nirwan Ansari (eds), *Neural Networks in Telecommunications*, Boston, MA: Kluwer Academic Publishers, 1995, pp. 271–85.

53 Jackel et al., op. cit. (52), p. 271.

and ‘successful applications’ were blurred as the group applied technology tested under narrow conditions at scale. Early applications appeared internally at AT&T in a document retrieval system that enabled users to browse the company’s technical publications. A second system automated processing of new service orders for parts of the AT&T network, while a third translated large volumes of densely printed tabular text from scanned documents into a structured ASCII format.⁵⁴ Success amidst controlled conditions thus triggered carefully managed trials that took place under the AT&T corporate umbrella, which, in turn, inferred the applicability of computer vision systems necessary for commercial deployments further afield.

Epitomized by a 1993 paper authored by Hans Graf detailing a formal partnership between Bell Labs and the US Postal Service, commercial deployments of character recognition systems remained reliant on the creation of roles that corrected system errors. The research, which followed a call in 1992 from the Postal Service for commercial partners, aimed to reduce the cost of hand-counting mail – a process estimated to cost forty dollars per thousand letters counted.⁵⁵ Announcing early results from the partnership, the paper claimed an accuracy of 98 per cent in ‘cleaner’ images, falling to 75 per cent in ‘noisy’ images.⁵⁶ Bell Labs researcher Craig Nohl, however, told the *San Francisco Chronicle* in 1993 that in practice the system could successfully recognize approximately three-quarters of the handwritten ZIP codes submitted to it. ‘The handwriting on the other 25 percent is so sloppy that the computer doesn’t try. It sends them to a human for sorting’, Nohl explained.⁵⁷ Bell Labs public-relations executive Robert Ford, speaking to *Business for Central New Jersey* in 1992 about the company’s partnership with the US Postal Service, offered that ‘this high tech research will relieve people of the drudgery of work and mundane tasks. The Information Age will free people to be more creative’.⁵⁸ Mara Mills has historicized the ‘essential role of human transcribers’ in correcting electronic outputs, while the clerical roles on which the performance of artificial neural networks depended have most recently been characterized as ‘ghost work’.⁵⁹ Indeed, the interventions by Ford and Nohl underline the central role of human overseers to the success of the group’s commercial computer vision applications. The provision of this safety net rendered invisible the very labour on which the success of industrial deployments relied, and in doing so the nature of the group’s networks as assemblages of data, institutional expertise, technical practice and clerical labour.

Commercial applications underpinned by technical labour intersected with increases in speed and performance as researchers tightened their grip on the language of neurophysiology. A 1992 paper examining the application of the ANNA chip introduced a year earlier to handwritten character recognition boasted a ‘speed advantage of 50 to 500 over conventional hardware’, which enabled the ‘neural network’ chip to process a thousand characters per second.⁶⁰ Increases in processing speed were described with

⁵⁴ Jackel *et al.*, *op. cit.* (52).

⁵⁵ Anon., ‘Leading companies bid for US Postal Service’s handwriting recognition contract’, *Computergram International Apt Data Services*, 1 July 1992, Ltd. No. 1954.

⁵⁶ Hans Graf and Eric Cosatto, ‘Address block location with a neural net system’, in Jack D. Cowan (ed.), *Proceedings of the 6th International Conference on Neural Information Processing Systems*, San Francisco: Morgan Kaufmann Publishers, 1993, pp. 785–92, 785.

⁵⁷ Anon., ‘Where high-tech miracles are born: research Mecca AT&T Bell Labs is targeting the marketplace’, *San Francisco Chronicle*, 28 June 1993.

⁵⁸ Anon., ‘Companies large and small in communications and biotech’, *Business for Central New Jersey*, 27 May 1992, p. 19.

⁵⁹ Mills, *op. cit.* (1); Mary L. Gray and Siddharth Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*, Boston, MA: Houghton Mifflin Harcourt, 2019.

⁶⁰ Eduard Säcker, Bernhard Boser, Jane Bromley, Yann LeCun and Lawrence Jackel, ‘Application of the ANNA neural network chip to high-speed character recognition’, *IEEE Transaction on Neural Networks* (1992) 3(2), pp. 498–505.

the language of neuroscience in 1993, when a *San Francisco Chronicle* profile of Bell Labs forecast a 'ten thousand-fold increase in the intelligence of microelectronic chips until they begin to approach the complexity of the human brain'.⁶¹ Highly reminiscent of the group's 1986 predictions concerning the future of conventional computing, both the research into artificial neural-network hardware and supportive comments in the media surface a through-line between scalability and generalizability. Implicit in this assumption is that replicating 'intelligence' in artificial neural networks is a matter of engineering; purpose-built hardware for artificial neural networks need only continue to develop along its existing trajectory in order to approach the sophistication of the human brain. The dynamic recalls the way in which AI researchers have consistently, as science and technology researcher Alison Adam has noted, extrapolated from a 'bounded problem solving situation to make an important claim about the nature of general problem solving'.⁶² Press speculation surrounding such projections was exacerbated by reports of carefully chosen applications that, as we have seen, took place within the AT&T group or with the essential support of workers to correct errors.

Conclusion

This paper has sought to disentangle the web of social and material contingencies that made up artificial neural networks designed and built by Bell Labs: the data sets on which the systems were trained and tested, the support provided by the organization and its AT&T parent, and the technical and clerical labour on which the reliable functioning of the systems was dependent. It has engaged the computational metaphor in neuroscience, which involves drawing parallels between the functioning of computers and the brain, in order to demonstrate the power of brain-computer imaginaries in promoting ideals of self-sufficiency and competency at the expense of the individual elements of the system. My core contention has been that the discourse of neuroscience enabled the systems to be cast as effective, reliable and autonomous entities capable of independent acts of recognition. Crucial for making this case is the 'proxy', which acts as the necessary form of make-believe that enables a sort of knowledge production reliant on the people, artefacts, places and moments invested with the authority to represent the world.⁶³ A critical component of the networks, the data sets used for the twin purposes of development and assessment, acted as a bridge between the practicality of conducting research and the representation of an extensive catalogue of possible forms and permutations of handwritten digits. It is this quality that permits ZIP codes to stand in for the base ten numeral system, electronic feature detectors to stand in for their organic counterparts, and artificial neural networks to stand in for the human brain.

For just as the data sets central to the functioning of the system enabled the representation of fragments of the world outside the group's laboratories, so too can the networks in which they exist be viewed as proxies for a broader swathe of complex phenomena, structures and systems. The power delegated to the networks as proxies, and the imaginaries surrounding them that stressed a likeness with the human brain, enabled conceptions of recognition and perception to be collapsed into a highly constrained sphere in which the systems rivalled human performance. In this way, the group's artificial neural networks – and commercial applications in partnership with AT&T or with the essential support of workers to correct errors – electrified the generation of brain-computer imaginaries through which the indispensable role of technical and clerical labour,

61 Anon., op. cit. (57).

62 Alison Adam, *Artificial Knowing: Gender and the Thinking Machine*, London: Routledge, 2006, p. 30.

63 Mulvin, op. cit. (6), p. 4.

institutional power and data sets used to train and test were made invisible. Through moves such as these the essence of the ‘neural’ network as an extensive technological system was eclipsed by persuasive portrayals that heralded a class of algorithm deemed to be worthy of the name.

Acknowledgements. With thanks to Richard Staley, Jonnie Penn, Aaron Mendon-Plasek, Matthew Jones, Dylan Mulvin, Stephanie Dick and Michael Castelle for comments and conversations that informed this work.