Pervasive Data Science

Technology is increasingly enabling us to instrument our physical environment with complex sensors and actuators, creating a connected world that generates huge volumes of data. This article presents opportunities and challenges in this new field of pervasive data science.

Nigel Davies
Lancaster University

Sarah Clinch
University of Manchester

Recent years have seen the rise of data as a central tenant in computing applications, products, research, and innovation, leading commentators to identify this rise as a big data "paradigm shift."[1] There has also been exponential growth in researchers' use of the term "data science" to describe the interdisciplinary field of collecting, drawing inference from, and acting on data. According to Google Trends, interest in search volume for the term has increased tenfold since 2010 (https://trends.google.com/trends). Beyond the hype, it is clear to see how our world is genuinely becoming one that is increasingly data centric, in which both physical and electronic services depend on the collection, analysis, and application of large volumes of heterogeneous data.

Examples of the success of data science abound. Applying new machine-learning techniques to problems such as speech recognition has made commonplace levels of performance that would have seemed impossible a few years ago. The transformative nature of this innovation can be seen in applications such as Apple's Siri, which reportedly receives over 2 billion requests per week.[2] These and other success stories have encouraged substantial levels of new research into algorithms and systems for data processing, storage, and visualization.

In contrast to this focus on data, pervasive computing has historically focused more on the user experience, principally motivated by Mark Weiser's original papers and his compelling vision of "calm computing."[3] Of course, data has often played a role in providing this user experience, and research areas such as activity recognition and location technologies are highly quantitative in nature. However, in recent years, the field of pervasive computing has invested significant effort into understanding the societal implications and applications of Weiser's vision. This focus on the human elements of pervasive computing is in contrast to the technology and application trends increasingly enabling us to instrument our physical environment with complex sensors and actuators and creating a connected world that generates huge volumes of interconnected data. The importance of these trends can be seen in the growing momentum of exemplars such as the Internet of Things (IoT), smart environments and smart cities. These applications demand a new focus on how we capture, process, and use data in pervasive environments.

In this article, we identify a new field of research that we call *pervasive data science*. Our objective is to highlight the importance of work in this area and identify examples of key research challenges for the community, thereby acting as a catalyst for new research. We define the field as follows:

> *Pervasive data science is research that exists at the intersection of pervasive computing and data science, characterized by a focus on the collection, analysis (inference), and use of data (actuation) in pursuit of the vision of ubiquitous computing.*

Example topics that we view as part of the pervasive data science agenda include new IoT

Published by the IEEE CS ■ 1536-1268/17/$33.00 © 2017 IEEE

sensor networks; new architectures for data processing at the edge of the cloud; algorithms for processing pervasive sensor data; new techniques for data visualization in pervasive environments; and cross-cutting concerns, such as privacy and trust. These topics highlight the inherently multidisciplinary nature of the field of pervasive data science. We're not simply talking about a shift toward more quantitative ubicomp research; rather, we're talking about a fundamental shift in which we view data as a key element in delivering Weiser's vision.

Example Application Areas

To what extent does pervasive data science really have an important role to play in our everyday lives? Understanding the breadth of data-driven pervasive applications is a worthwhile foundation for defining the field. In the following examples, we illustrate some of the potential opportunities for data-driven pervasive applications. Although the application areas might not be completely new, to date, pervasive computing has failed to make them an everyday reality. The new capabilities provided by pervasive data science could be transformative in enabling such applications.

Augmented Cognition

Our everyday lives feature a huge range of cognitive activities, and technologies and artefacts have long played a role in supporting these—we write lists to evaluate the pros and cons of an important decision, take photographs of experiences we want to remember, and record voice memos to help with productivity. As data collection, processing, and presentation become all-encompassing, the potential to more effectively support these processes increases dramatically. Indeed, many have proclaimed that industrialization and technology have greatly increased the cognitive demands on human beings,[4] so it seems only fitting that technology would now help us address these concerns by supporting cognition and reducing our cognitive load. The exact nature of these technologies in terms of user interfacing is currently unclear; nevertheless, there is undoubtedly an important role for data-intensive pervasive systems in a huge array of cognitive processes, including decision making, risk evaluation, mood and emotion regulation, creative thinking, attention and information processing, and retrospective and prospective memory.

Designing these systems will require a clear understanding of cognition. Psychologists have recently started distinguishing between two metaphorical systems of thought: System 1 ("intuition") and System 2 ("reasoning").[5] While operations of System 1 are fast, automatic, and effortless, they are also often emotionally charged, governed by biases and habit, associative, and difficult to control. By contrast, the operations of System 2 are slower, effortful, resource-intensive, and deliberately controlled, but they are also more adaptive and neutral with regard to emotion and bias. Pioneered as a metaphor for psychologists and behavioral economists, this distinction might also be valuable in understanding target processes that support pervasive data applications.

Critically, System 1 thinking is often used to explain flaws in our cognition—such as in the evaluation of risk or in decision making—because we have either insufficient time or inclination to engage in deeper processing. If technology can reduce the barriers to System 2 thinking or can provide such thinking in System 1 timeframes, then many of the impairments to reasoning caused by System 1's inherent biases could be overcome. For example, perhaps data-processing techniques could be used to reduce the effort of System 2 thinking, while the additional data provided by pervasive sensing could provide richer input for such processing.

> Pervasive data science could help address growing concerns with regard to mental health problems and cognitive disorders.

In addition to helping address flaws in cognition, pervasive data science could help address growing concerns with regard to mental health problems and cognitive disorders. In 2015, the US National Institute of Mental Health (www.nimh.nih.gov) estimated the prevalence of a diagnosable mental disorder at 1 in 4 adults, with nearly 1 in 25 having serious functional impairment due to a mental illness.[6] In terms of cost, it is estimated that this amounts to more than US$300 billion per year. Similarly, the rising prevalence of dementia incurs costs of around $200 billion annually, more than both heart disease and cancer, largely due to the need to provide institutional and home-based long-term care for individuals who can no longer perform daily activities without support. Providing unobtrusive support that helps individuals maintain their own identities and manage their moods, emotions, attention, memories, and thought processes could be critical in reducing the costs associated with cognition and mental-health-related illnesses and decline.

Autonomous Vehicles

Ever since Weiser's original description of the "foreview mirror,"[3] pervasive computing has looked to enable smart transportation, and autonomous vehicles represent a natural extension of this work. For many years, computation and electronics have increasingly been added as features of cars and other vehicles. However, the wealth of

available sensor data (both in the environment and on the vehicle), together with advances in machine learning to interpret the data and predict future environmental changes, means that we have now moved from automation to genuinely autonomous vehicles. The arguments for such vehicles are well rehearsed—improved safety is perhaps the most obvious, along with eco-driving and a reduced burden for users.

Although one could simply envision autonomous vehicles as enhanced cars that improve each individual journey by allowing those being transported to travel more comfortably, safely, or economically, the reality is that autonomous vehicles will offer new opportunities for transportation and logistics as a whole—forming part of a new vision of "mobility as a service." This vision extends beyond autonomous vehicles and imagines a world in which mobile applications let travelers dynamically select the most appropriate form of transportation to achieve their mobility objectives, with potentially profound implications for transport infrastructure, urban planning, and economics.

Smart Spaces

Exemplars such as IoT, smart environments, and smart cities are arguably the most obvious current applications of pervasive data science. Technology increasingly lets us instrument our environment with sensors and actuators, creating a connected world that generates huge volumes of complex data. Of course, the addition of "smart" into our environments is not new for pervasive computing. Weiser's ubicomp vision included domestic appliances that interpret future needs based on recent activity, neighborhoods that track mobility patterns, and an office that supports awareness and communication between remote colleagues. However, progress in data science is beginning to provide tools with which to realize elements of this vision and to offer new services that go beyond the original vision based on the collection, analysis, and application of data in our physical environments.

Applying pervasive data science to our physical environments offers a wealth of opportunities for improving quality of life through access to smarter and more appropriate services. While some of these advantages span almost all of our physical spaces, there are also specific goals that can be addressed in particular target environments. For example, we can use pervasive data to make our workspaces more pleasant and efficient by extrapolating from the data which physical conditions lead to a satisfied and productive workforce, how we can best foster inter-team collaboration, and which portion of a workflow offers the most potential for optimization. Schools and places of education can build an understanding of how nonclassroom environments (such as corridors and social spaces) can be used to complement learning. Similarly, using pervasive data science in outdoor environments can help tackle challenges such as climate change, traffic congestion, and urban air pollution.

While many of these examples approach smart spaces as something that benefits populations and organizations as a whole, pervasive data science also lends itself perfectly to the tailoring of physical environments to provide a personalized experience for the individual. Shared spaces could be uniquely configured to respond to the user, changing not only the aesthetics but also the physical configuration of the space itself.

Challenges

In considering the challenges associated with pervasive data science, it's worth reflecting on the challenges traditionally associated with data science and pervasive computing. Although data science is a relatively new field, it draws on many years of prior work in areas such as statistics, algorithms, and databases. Data science challenges are often based on the characteristics of big data, as expressed in the three "Vs"—volume, variety, and velocity[7]—and sometimes supplemented with an additional V in the form of veracity (we revisit these challenges later), but it's worth noting that general papers on the challenges of data science are rare.

By contrast, there are a large number of papers that have articulated challenges relating to mobile and ubiquitous computing, including Mahadev Satyanarayanan's "Fundamental Challenges in Mobile Computing"[8] and Weiser's original paper.[3] Indeed, the inaugural issue of *IEEE Pervasive Computing* in 2002 focused on articulating the challenges that still existed in the field 10 years after its inception. Many of these challenges remain and cover a broad space, including the development of appropriate systems architectures, new forms of user interaction, and cross-cutting concerns such as ease of deployment and system maintainability.

Here, we present some of the new challenges that arise at the intersection of these two fields. We structure our discussion in terms of the key stages of data processing: data collection, inference, and subsequent action.

Data Collection

The first stage in any data pipeline is data collection (and cleaning). Pervasive data can originate from a wide range of sources, including sensors embedded in the environment, sensors attached to users, and explicit user input (as in the case of initiatives such as citizen science). Although the design of any large-scale data collection system is nontrivial, the emergence of systems to support pervasive data science gives rise to some exciting new research challenges.

How do we manage complex models of data ownership? In conventional sensor systems, the question of data ownership is relatively clear—the data owner is normally the same as the owner of the system being instrumented. Similarly, most data scientists assume data ownership has been resolved prior to them obtaining the data for analysis.

However, in a world of pervasive sensors, the question of data ownership becomes significantly more complex. For example, in many smart environment applications, the same space might be instrumented by many different stakeholders. Mobile users might bring their own sensors, or wish to use those of their peers, and the use of spaces and sensors could be highly transient. As a result, the ownership of any given data stream (or combination thereof) might be unclear.

Can we develop techniques for automatically resolving data ownership in such pervasive computing scenarios? How do we model shared ownership of data? How do we accommodate ownership expectations when considering personal data that users typically perceive as belonging to them, even though they might not own the sensing infrastructure (such as information on energy and activities in a domestic environment collected by a smart heating management system)? How do we ensure that ownership of data reflects transient use of pervasive sensors and spaces?

Can data provenance be ensured in pervasive systems? Determining the provenance of data in existing systems is a well-documented research challenge, with solutions typically involving techniques such as audit trails based on digital signatures. However, in pervasive systems, many new aspects of data provenance become important.[9] For example, how do we capture the identity and motivations of humans involved in sensor placement, given that even a small bias in placement can significantly influence the data captured? Once provenance data has been captured, how should this be presented to users to enable them to understand the likely impact on the data? Given that pervasive systems can involve complex data pipelines, it makes sense to record the details of these pipelines, and yet some aspects (such as data redaction policies) might be sensitive. How do we balance the need for end-to-end provenance with the need to mask the identity and motivations of some users?

How do we balance the insatiable demand for data with privacy concerns? Privacy has long been recognized as a challenge in pervasive environments.[10] Despite extensive research, the challenge of protecting user privacy remains; indeed, it's becoming significantly more difficult to address as the number of sensors in the environment increases. Are we now reaching the point where the only way to effectively protect user privacy is to limit data collection? If so, new architectural solutions will be required to enable data to be quenched at the source. We, along with others, have proposed a model in which users can control the release of data from their homes.[11] While the model is simple—"users should be able to control the release of their own data"—the implementation is complex, necessitating the introduction of new architectural components, such as privacy mediators, that can denature data prior to disclosure.

How will data subjects provide informed consent? Related to the issue of user privacy is the challenge of supporting informed consent. How should we inform users of data collection in a pervasive data environment? It's clearly impractical to explicitly prompt users every time they enter an environment in which data about them is being captured. However, it's equally important to ensure that users understand and consent to the collection of information and, crucially, have a genuine option to opt out of data collection.

Once consent has been provided, many systems provide tools for data collectors to track subject consent. However, there is a dearth of tools that enable users themselves to track when and where they provided consent. This opens up the opportunity for unscrupulous data collectors to simply claim that consent has been provided—how many users could really assert with confidence whether such a claim was true?

Although this discussion suggests that pervasive data science is likely to make managing consent extremely challenging, pervasive data science also offers the potential for entirely new ways of managing consent, based on systems that automatically learn a user's preferences and behaviors and infer whether or not to automatically provide consent.

Related to the issue of user privacy is the challenge of supporting informed consent. How should we inform users of data collection in a pervasive data environment?

Inference

Inference lies at the heart of what many consider to be data science and represents the process of analyzing data to gain understanding and insight.

How does pervasive computing affect traditional data science challenges? In future pervasive environments, the volume of data is likely to dwarf that produced by most existing data systems. For example, while classic data systems can examine feeds such as web browsing histories or social media posts, pervasive data applications operate in a world in which sensors capture every aspect of a user's experience. As a point of reference, current life-logging cameras capture over 2,000 images per day, far exceeding the number any user might manually process or post in

a typical day. How can we store and process such volumes of data?

Widespread user and environmental sensing is also likely to lead to a variety of data previously unseen, creating heterogeneous datasets in terms of format, frequency, and quality (among others). This raises challenges in terms of data consistency but opens up exciting new possibilities, such as effectively combining data from a wide range of sensors. Although one individual sensor is unlikely to produce high-velocity data, a combination of sensors in any given environment could lead to data aggregators experiencing streaming data at an unprecedented velocity. This raises important questions regarding future data-processing architectures.

How should we architect pervasive data science environments? Existing data systems tend to assume that large-scale data centers are available to carry out the processing necessary to draw inference. When sensors are the source of this data, they are typically assumed to be "dumb" sensors, with all of the processing being conducted in the cloud. However, research has clearly demonstrated the shortcomings of a purely cloud-based model,[12] and new architectures have been proposed that provide data processing at the edge of the cloud.

What are the correct processing architectures for future pervasive data environments? Such environments might offer a wide range of options for hosting data processing—from highly sophisticated sensors that can carry out significant levels of analysis on-board, through edge-of-cloud solutions, to micro and full data centers. To what extent does the exact configuration of processing elements used depend on the intended applications, and how will it be influenced by security and privacy concerns?

Actuation

Traditional data science has often focused on information presentation as its key output. Likewise, other fields that work with large emerging datasets (such as computational social science[13]) reach their end with the analysis of data representing human behavior patterns. Each of these fields achieves actuation only through humans implementing changes based on the outcome of data. By contrast, actuation has been a common feature of ubiquitous computing since Weiser's early vision, and exemplars such as IoT and smart cities continue to demonstrate its importance. Although pervasive data science requires inference to deliver value, its data lifecycle doesn't end with analysis; instead, pervasive data applications offer opportunities and challenges in both new forms of information presentation and physical actuation.

How can we best use pervasive technology to visualize rich datasets? Data scientists have always sought new ways of visualizing data. Pervasiveness provides a wealth of new presentation opportunities, enabling the process of engaging with data to be switched from one of active interpretation to one of passive immersion. Weiser observed that natural environments can convey a wealth of information that can be readily absorbed and yet still deliver a positive user experience, and several trends in pervasive displays are helping to realize a similar paradigm for digital interactions.[14]

Given the range of available technologies, how do we select the most appropriate medium for engaging users? How do we take into account factors including the contextual relevance of the device, the scale and resolution of data that can be represented, the shared or private nature of the content, and individual aspects such as attention and task engagement? While data science has focused primarily on visualizing data for expert users, pervasive data potentially creates a requirement for information presentation to become accessible to a wide range of individuals. This "data-for-the-masses" poses considerable new challenges. How do we develop a set of patterns for comprehendible representations (not necessarily visual) that are accessible to populations of different ages, education levels, and cultural backgrounds?

What new forms of data-driven actuation will emerge? Visualization is just one output form for pervasive data, and numerous other forms of data-driven actuation exist, including data-driven control and adjustment of our environments. Such adjustments might take the form of a slow transition that optimizes the performance of a space, or the adjustments could involve a more dynamic process that personalizes an environment based on its changing purpose and the individuals within it. What new possibilities for smart spaces will emerge with widespread data availability? How will users be made aware of such data use, and how will they exercise control over actuation? Of course, data-driven smart spaces are just one example, and numerous other forms of data-driven actuation are likely to emerge in areas such as IoT.

Pervasive Data Science in Context

The questions we've raised here highlight the research opportunities that exist at the intersection of data science and pervasive computing. Pervasive computing brings new challenges to conventional data science tools by virtue of both the scale at which it operates and its focus on the relationship between technology and users. However, the successful application of data-driven approaches to pervasive computing challenges open up the possibility of genuinely smart environments and applications, as described earlier.

Of course, there have been numerous attempts to transfer the techniques and insights of data science to a broad set of disciplines. For example, almost 10 years ago, David Lazer and his colleagues described the emergence of data-driven computational social

science—an application of data at scale to describe individual and group interactions.[13] Likewise, social and community intelligence (as articulated by Daqing Zhang, Bin Guo, and Zhiwen Yu in 2011[15]) leverages these same tools to reveal human behavior patterns and dynamics. Furthermore, ubiquitous computing itself has often had a data dimension (consider the computational location applications of John Krumm and others[16]), incorporating qualitative and quantitative data from a range of sources to capture both "the masses" and the individual. While domains such as computational social science and quantitative ubiquitous computing have focused predominantly on data collection and analysis, one unique feature of pervasive data science is its end-to-end use of data—not just collection and analysis but also actuation. Indeed, we believe that the emerging trend for pervasive data science is the natural progression of such disciplines, but broadened to reflect the original ambition of Weiser's ubiquitous computing.

Another key distinction of pervasive data science is its interdisciplinary nature. While a field such as computational social science integrates computing methods with those of the social sciences, pervasive data science must bring together social sciences and humanities with a broad set of computer science expertise—including theory, systems architecture, and human-computer interaction. The challenges we have highlighted all span multiple fields of computer science (see Figure 1), which is deliberate. This intersection offers new opportunities for interdisciplinary research and development and is a distinct feature of this emerging field.

	Theory	Systems	People
Collection	Provenance and ownership		
	Algorithms for signing streaming data that are optimized for use on low-power sensors.	End-to-end secure architectures and protocols for high-velocity data streams.	Techniques for presenting provenance to users. Models of data ownership.
	Privacy and consent		
	New algorithms for data denaturing.	System architectures that encompass privacy by design.	New UI techniques for obtaining informed consent.
Inference	Challenges to traditional data science		
	Techniques for compensating for highly variable and low-quality data.	Storage systems for high-volume, high-velocity data.	Tools for helping users understand inferences drawn from their data.
	New architectures		
	Algorithms optimized for distributed operation at the edge of the cloud.	Novel architectures to support computational off-loading.	Acceptable business models for data processing and service provision.
Actuation	Pervasive technology for information visualization		
	New algorithms for info-viz on distributed displays.	Architectures for display coordination.	Guidelines for preventing information overload in pervasive environments.
	New forms of data-driven actuation		
	Formal models of how environments respond to different data inputs.	Protocols for secure communication with IoT actuators.	Interfaces for user control of data use.

Figure 1. Examples of emerging challenges in pervasive data science. Six core challenge areas span multiple fields of computer science and present new opportunities for interdisciplinary activity.

Early Case Studies

As a prime example of these interdisciplinary challenges in practice, we return to our earlier scenarios. We consider two case studies of work in the area of augmented cognition—specifically, task assistance and augmented memory. These examples provide an important illustration of the end-to-end nature of pervasive data science, while providing a foundation from which challenge areas can be considered and addressed.

Cognitive Assistance

Researchers at Carnegie Mellon University have set out a vision for "angel on your shoulder" cognitive assistance that takes the familiar models of detailed directions given by current GPS systems and applies them to everyday living.[15] Such systems combine wearable sensing and presentation with local processing to guide users through complex tasks, telling them what to do next (with audiovisual cues) and correcting any erroneous actions. While arguably an ambitious concept, with very broad parameters that apply in virtually all facets of everyday life, the CMU-based group has focused predominantly on support for well-defined tasks that might require the prompt application of specific knowledge or skills (such as administering first aid or engaging in a competitive sport). A key challenge is that completing such tasks is both latency sensitive (a decision must be made in finite time) and resource intensive (requiring processing of the current context, identification of relevant skills or knowledge, and determination of the correct behavioral response).

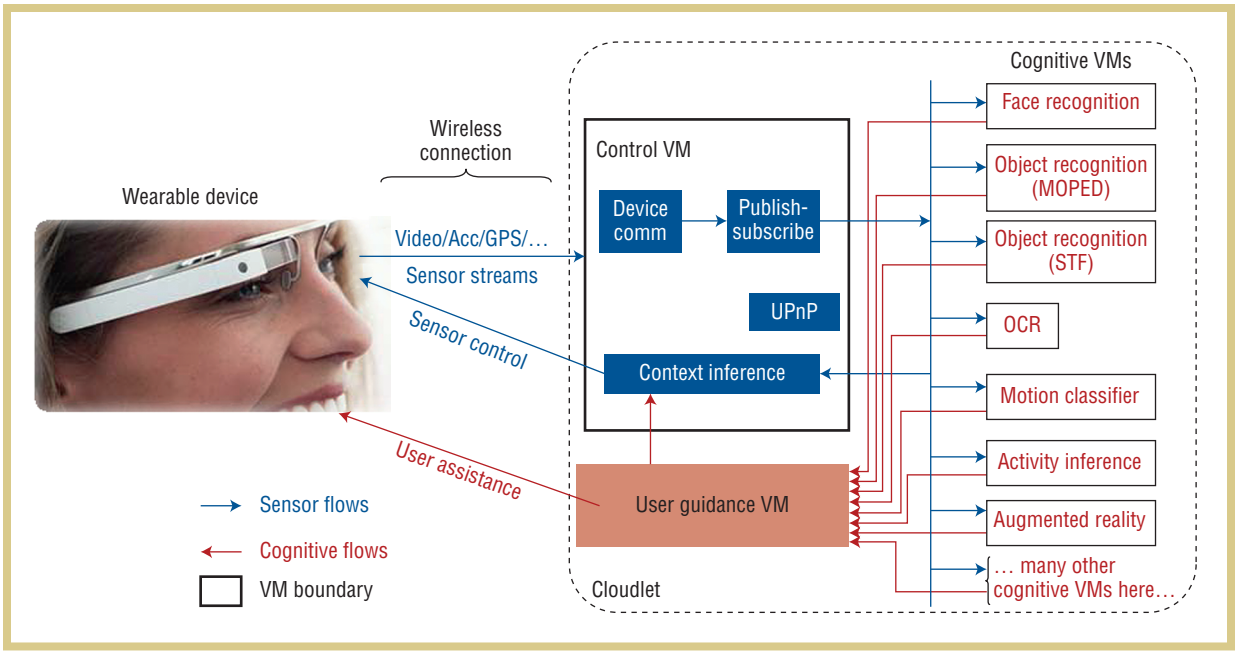

Figure 2. The Gabriel architecture for cognitive assistance.[17] Cognitive processes are encapsulated in distinct virtual machines—the output of each process is integrated by a control VM and then delivered back to the user through their wearable device. (Source: Mahadev Satyanarayanan; used with permission.)

To address these challenges, the CMU group created the Gabriel platform,[17] which uses a cloudlet computing infrastructure[10] that encapsulates each cognitive process (face recognition or motion classification, for example) in its own virtual machine (see Figure 2). Each VM can then independently process incoming sensor data from a control VM. Any resulting output from the cognitive processing is then passed back to a shared user guidance VM, which can aggregate the results and perform any high-level integration or further processing needed to deliver assistive input to the user. The Gabriel architecture has been verified with an initial set of demonstrators, indicating that cloudlets could be a key part of future cognitive assistance applications.

The assistive output provided in Gabriel and its associated demonstrators is a clear illustration of pervasive data science in practice. Each of the four demonstrator applications relies heavily on worn sensors to capture rich (typically visual) data feeds representing the task to be completed and the associated context. Computer vision and other data processing techniques then interpret sensor input and predict the outcome of multiple future behaviors, leading to an optimal target behavior that the assistant should support the user in achieving. Further processing of historical data can then identify the best cues and presentation medium for this user in the given context. Finally, the system presents input to the user and continues collecting data to close the loop and measure the ongoing success of its advice.

Human Memory Augmentation

Our own research addresses human cognition in terms of retrospective and prospective memory. Human memory is critical to self-identity and to the success of most of our everyday activities. While prospective memory (memory for intention) is a relatively constrained problem space, retrospective memory encompasses recollection of personal experiences (episodic memory), knowledge acquired (semantic memory), and motor skills learned (procedural memory). With this in mind, we focus here predominantly on episodic memories (although many of our approaches might well generalize beyond these).

Through a series of prototypes and deployments, our research has led us to an architecture that combines mobile, wearable, and environmental devices to capture a rich representation of an individual's experience (and thus derived knowledge),[18] including occurrences in which the human memory itself was known to have failed. These data streams serve as input to a graph-based storage, the *memory vault*, which attempts to reflect human cognition in its model of how individual pieces of information connect together to form a single digital memory, and also its model of the interconnection between multiple digital memories. Because many human experiences are the product of social interactions, and because memory of those experiences is continuously shaped by ongoing interaction with those (and

other) individuals, a single centralized digital memory can't meet an individual's needs. Instead, we anticipate that our memory vault will be a distributed memory, with interconnections spanning multiple instances and with data from others being made readily available to those who shared an experience.[19]

Human memories continuously draw on traces formed over many years, and ongoing processing and inference are critical to the success of memory augmentation systems. Unlike the example of task assistance, which focused primarily on real-time processing, our memory augmentation architecture heavily uses continuous long-term data aggregation and inference, and the output of these is stored alongside raw experience data in the memory vault. When an individual then engages in a situation that would be facilitated by access to extended memory, our architecture can pass current contextual feeds to the memory vault and select an appropriate cue for presentation. The system then presents input to the user and continues collecting data to close the loop and measure its ongoing success.

In line with challenges outlined earlier, our research in the area of augmented memory has highlighted significant security and privacy concerns related to the acquisition, processing, and presentation of datasets.[18] Bystander privacy is one issue often raised (and is thus the target of considerable research to date[20,21]), but many more subtle data challenges also emerge. For example, the issue of provenance discussed earlier becomes important in ensuring that digital memory is an accurate representation of what occurred, particularly since the very need for augmented memory means the human might be unable to determine this. The phenomenon of recall-induced forgetting is a good example of a threat that arises only when our understanding of human cognition is applied to the design of data-intensive pervasive applications—in this case, psychology theory indicates that rehearsal of one memory has a detrimental effect on the subsequent recall of related memories, a clear concern for those designing technology interventions in this space.

Recent developments in fields such as machine learning have demonstrated the potential of data science to transform our ability to deliver solutions to traditionally very demanding problems, such as speech recognition. These developments, coupled with widespread deployment of sensing and actuating technologies, mean that there is the potential for pervasive computing to adopt an increasingly data-driven approach—one of pervasive data science.

In analyzing the challenges that such an approach raises, we observe that many are cross-cutting in two distinct axes. First, the challenges span all stages of the data pipeline—from data collection through to data visualization and actuation. Second, the challenges require multidisciplinary approaches, because they raise issues in terms of systems, algorithms, and people. For example, protecting user privacy in future pervasive environments demands that new systems work on topics such as privacy mediation, new algorithms for efficient data denaturing, and new techniques to enable users to visualize and control the release of their personal data. However, the payoff for developing solutions to these problems is significant—applications such as augmented cognition and memory have the potential to transform our society, delivering benefits to millions.

Cross-cutting multidisciplinary problems are, of course, common in traditional pervasive computing research, so the field is particularly well-suited to address these types of problems. Pervasive computing has always favored researchers who adopt a holistic view, and the emergence of pervasive data science serves to reinforce the validity of this approach. ■

the AUTHORS

Nigel Davies is a professor of computer science at Lancaster University, where he is also codirector of the Data Science Institute. His research is in the area of pervasive computing, including systems support for new forms of data capture and interaction, and his work is characterized by an experimental approach involving large-scale deployments of novel systems with users. Contact him at n.a.davies@lancaster.ac.uk.

Sarah Clinch is a computer science researcher and lecturer at the University of Manchester, UK. Her research interests include applications for human cognition, pervasive display deployments, and privacy and personalization in ubiquitous computing systems. Clinch has a PhD in computer science from Lancaster University. She is an inaugural member of the ACM Future of Computing Academy. Contact her at sarah.clinch@manchester.ac.uk.

ACKNOWLEDGMENTS

This research is partially funded through the Future and Emerging Technologies (FET) program within the 7th Framework Program for Research of the European Commission, under FET grant number 612933 (RECALL) and the EPSRC under grant EP/N028228/1 (PACTMAN: Trust, Privacy and Consent in Future Pervasive Environments).

REFERENCES

1. R. Schutt and C. O'Neil, *Doing Data Science: Straight Talk from the Frontline*, O'Reilly Media, 2013.
2. J. Clover, "Tim Cook: 'We Don't Buy' the Need to Give Up Privacy for AI," *MacRumors*, 25 Oct. 2016; www.macrumors.com/2016/10/25/apple-artificial-intelligence-privacy.
3. M. Weiser, "The Computer for the 21st Century," *Scientific Am.*, vol. 265, no. 3, 1991, pp. 94–104.

4. C.S. Saunders and A.-F. Rutkowski, "Growing Pains with Information Overload," *Computer*, vol. 43, no. 6, 2010, p. 96.

5. D. Kahneman, "Maps of Bounded Rationality: Psychology for Behavioral Economics," *The Am. Economic Rev.*, vol. 93, no. 5, 2003, pp. 1449–1475.

6. *Brain Disorders: By the Numbers*, McGovern Inst. for Brain Research at MIT, Jan. 2014, https://mcgovern.mit.edu/brain-disorders/by-the-numbers.

7. D. Laney, *3D Data Management: Controlling Data Volume, Velocity and Variety*, tech. report, Gartner, 2001.

8. M. Satyanarayanan, "Fundamental Challenges in Mobile Computing," *Proc. 15th ACM Symp. Principles of Distributed Computing*, 1996, pp. 1–7.

9. B. Knowles, "Emerging Trust Implications of Data-Rich Systems," *IEEE Pervasive Computing*, vol. 15, no. 4, 2016, pp. 76–84.

10. M. Langheinrich, "A Privacy Awareness System for Ubiquitous Computing Environments," *Proc. 4th Int'l Conf. Ubiquitous Computing* (UbiComp), 2002, pp. 237–245.

11. N. Davies et al., "Privacy Mediators: Helping IoT Cross the Chasm," *Proc. 17th Int'l Workshop on Mobile Computing Systems and Applications* (HotMobile), 2016, pp. 39–44.

12. M. Satyanarayanan et al., "The Case for VM-Based Cloudlets in Mobile Computing," *IEEE Pervasive Computing*, vol. 8, no. 4, 2009, pp. 14–23.

13. D. Lazer et al., "Life in the Network: The Coming Age of Computational Social Science," *Science*, vol. 323, no. 5915, 2009, pp. 721–723.

14. S. Clinch, J. Alexander, and S. Gehring, "Pervasive Displays for Information Presentation," *IEEE Pervasive Computing*, vol. 15, no. 3, 2016, pp. 14–22.

15. D. Zhang, B. Guo, and Z. Yu, "The Emergence of Social and Community Intelligence," *Computer*, vol. 44, no. 7, 2011, pp. 21–28.

16. D. Delling et al., "Navigation Made Personal: Inferring Driving Preferences from GPS Traces," *Proc. 23rd SIGSPATIAL Int'l Conf. Advances in Geographic Information Systems*, 2015, article no. 31.

17. Z. Chen et al., "Early Implementation Experience with Wearable Cognitive Assistance Applications," *Proc. 2015 Workshop on Wearable Systems and Applications* (WearSys), 2015, pp. 33–38.

18. N. Davies et al., "Security and Privacy Implications of Pervasive Memory Augmentation," *IEEE Pervasive Computing*, vol. 14, no. 1, 2015, pp. 44–53.

19. A. Bexheti, M. Langheinrich, and S. Clinch, "Secure Personal Memory-Sharing with Co-Located People and Places," *Proc. 6th Int'l Conf. Internet of Things* (IoT), 2016, pp. 73–81.

20. P. Aditya et al., "I-Pic: A Platform for Privacy-Compliant Image Capture," *Proc. 14th Ann. Int'l Conf. Mobile Systems, Applications, and Services* (MobiSys), 2016, pp. 235–248.

21. M. Fan, A.T. Adams, and K.N. Truong, "Public Restroom Detection on Mobile Phone via Active Probing," *Proc. 2014 ACM Int'l Symp. Wearable Computers* (ISWC), 2014, pp. 27–34.