# Personal Archives, Privacy, and the Digital Afterlife

Dai Llewellyn Davies – Brindabella.id.au

**PA**: noun

1 Personal Archive: continuous life-long record of visual, audio, biometric and ambient data.

2 Personal Assistant: natural language interpreter for a Personal Archive with quizzing and command capabilities. Includes visual recognition.

3 Personal Avatar: arbitrary visual representation used to interface a Personal Assistant with the visual world. (for their physical form see **bot**)

The machine sitting in front of me is very versatile. It's based on a simple but powerful and general design – the Turing Machine. In its abstract form it was dubbed the Universal Turing Machine and formal logic proved it to be general in its capabilities. It has also proved to be flexible in practice. The downside of all this flexibility and generality is an inherent lack of security.

In this article I outline an alternative constrained architecture, the PA, that is designed with privacy and security as the principal challenges. This raises questions like: Why bother? What's all the fuss about privacy anyway? Surely if you haven't done anything wrong you've got nothing to hide? In the past, haven't people lived in small communities where privacy was unknown? Doesn't the present interest in social media show that many people want to share the details of their lives with others – even strangers? These are all valid questions and deserve answers. To answer them we need to put them in context – past, present and future.

We are told that automated surveillance of public places is justified as a means of reducing criminal activity. No doubt it does – where it's applied. No doubt, also, that it can just shift this activity to other settings. Cyberspace is the new, and global, playing field.

Staying with physical spaces for a moment we can see that things now are not the same as they were with old-fashioned CCTV. Closed Circuit TV was just that – closed. To access it meant physical access to videotapes and hours spent viewing them – if they hadn't been overwritten to save tape. Now cameras are networked and digitally accessed by pattern recognition algorithms looking at facial features, gait, mood analysis, and so on – and the records are usually permanent.

Now, many thousands of people have legal or illegal access to this information and a few individuals are inevitably drawn by money, mischief or malice to betray this trust. Of these three motivations, mischief may be the most insidious. Anyone for a game of BustUp? We each select a happily married couple, then triggering a family break up wins a point per family member. The further you are up the game status table the better the job offers.

People have played these games for millennia but now they can be anonymised, depersonalised and globalised – even automated. Do we reach a stage when simulation and reality mix to the point where you find you've been included as a character in a game without either you or the players knowing?

I could go on but I'm not out to scare – or ‘terrify’ as semantic inflation puts it these days. There's quite enough of that going on without me adding to it. I just want to make the point that there are issues to be considered and when some of us start to get a little queasy about the creepiness of it all, there

are technological alternatives – ones that also open up new opportunities.

It's true that many people in the past lived in small communities where privacy was minimal. Many people still do, but in this situation people know each other and the lack of privacy is usually reciprocal. Small communities also tend to be wary of strangers, particularly where policing is minimal or non-existent. Also, I expect that many people in such societies value highly what privacy they do have.

Our social media are an interesting experiment but one that people are still learning to adapt to. Announcing to the world that your cat has just been sick over the carpet might not have major consequences but dropping an indiscreet, or just ambiguous, comment about an employer, relative or friend when you're in an irritable mood can be a life changing act. A PA could provide discretion and accuracy checks on what we write before it goes public.

It's an issue that will become more important as people move into life-logging or full-life archiving, as many inevitably will. Why would we want to record everything we do? I don't have any great need for visual logging beyond what my phone can provide but I've longed for audio archiving for personal note taking – the ability to immediately record anything I say and, crucially, ready access into automatic transcripts. It could be a valuable prosthetic for those of us with poor or infuriatingly selective memory. None of us have perfect recall.

I think the main motivator that will push archiving into common practice will be medical. We each have a unique metabolism and react differently to foods, medications, different forms of stress, sleep cycles – even music or particular words. Large volumes of continuous (and hopefully

anonymized) biometric data will provide information that will revolutionise medicine, making it truly scientific. Simple blood pressure monitors and motion detectors are available now and lab-on-a-chip technologies are down-scaling towards a point that will allow detailed chemical composition of blood to be monitored continuously with a small implant.

Is a technical solution for PA security possible? I think so, but only partially with our current machine architectures. There are some basic design requirements for a secure and trusted PA architecture that can't be achieved with present architectures. A fundamental requirement is that people have complete and confident control over their interactions with the digital world. To achieve this the basic architectural principles need to be simple and readily understood by the average user.

*Memory should be Write-Once-Read-Many*: That the record be indelible is implied in the word ‘archive’. It is a permanent and exact record, even if intermittent, not a reconstructable history. This is a necessary requirement for the system to be trusted.

*Access should be restricted to a gatekeeper module*: The architecture should not provide any physical means of reading the archive other than through a hardware gatekeeper. Access via the gatekeeper should be under the sole, instantaneous control of the owner of the PA. This is a necessary condition for privacy.

*It should record every action it performs*: That the gatekeeper record *all* its actions as part of the archive is a necessary requirement for reconstruction, analysis and verification of its actions. This underpins both trust and privacy.

*The control logic should be expressed in natural language*: The operational rules don't need to be translated into a low level computer language. A human language should be *the* operational language of the device down to the hardware level. This gives operational transparency and a direct means for the owner to provide instructions and check that they are being interpreted correctly.

*There should be a core set of standard access rules*: These would provide trusted answers to basic questions such as 'Who are you?' along with diagnostic evidence that the answer was derived from the core rules. The core should use an unambiguously defined subset of the natural language in use.

*Manufacture should be completely transparent*: If you are going to trust this device you need to know what's going on inside it, or rely on a wide community of users who have checked the system you start with. This is probably the most difficult requirement to satisfy since it relies on trusting others with the construction. The only way I can see this being achieved is through multiple open source projects with a diverse range of people constructing the units.

Currently, the missing ingredient for the full development of PA technology is robust automated speech recognition (ASR) that will make reliable transcription and indexing of speech possible. Unfortunately, research into ASR is stuck in a rut. The need for a strategic retreat from current approaches was recognised as far back as 1996 when it was discussed extensively in the ASR literature (e.g. Bourlard et.al., *Towards increasing speech recognition error rates*) but little seems to have changed since then. I think an open web-based project could run rings around current academic and industry efforts.

The time is ripe for web-based scientific projects. Little, if any, money is needed – just imagination, creativity, and sustained effort that is not systematically constrained or derailed by the publish-or-perish imperative of academia.

There are several big advantages to be gained from using natural language in a PA. The most obvious is that most of us are fluent in at least one. This allows us to instruct a PA without the need for a specialist programmer. English has an advantage over current computer languages in that it combines procedural, declarative and interrogative functions. By this I simply mean that an English sentence can be an instruction, a statement, or a question.

With computer languages these functions are handled by separate languages such as Javascript (a procedural language consisting of a sequence of low-level commands), HTML (HyperText Markup Language for defining web page structure) or SQL (a Structured Query Language used in extracting information from database tables). Integrating these can be quite messy and each imposes strict constraints on the type and structure of the processed information. The expressive power of English can cope much better with 'found data' in its original form – likely to be English itself.

To tap into the meaning of sentences and interrelate their meanings we need a parser to resolve their structure, and an inference engine to trace the logical and semantic relationships between them. Both these technologies have advanced to the point where they can be user-modified and extended to tailor them to a particular user's requirements – just the kind of thing that's best done locally by you rather than some centralised algorithm trying to second-guess you.

The difficult side of English is its ambiguity. Most English words have more than one meaning. A ten word sentence where each word has two possible meanings gives over a thousand possible alternate interpretations. Fortunately, most of these interpretations can be dismissed as ungrammatical. The remainder must be evaluated, as our brains do, using the context of the sentence to constrain meaning.

This is an unavoidable, and often serious, problem with found data but it can be circumvented. By instructing a PA interactively the user can get immediate feedback and specify the required meaning or alter the wording of a sentence, gradually building a knowledgebase of personal usage and preferences. The alternative is centralised systems thrashing about trying to find ways of dealing with millions of different people and inevitably coercing them into accepting a generalised compromise – a technologically generated hive mind curbing the fundamental requirement of creativity and cultural evolution – diversity.

With a command capability that goes well beyond asking data-access queries – something we can instruct to perform tasks for us – we've moved from Personal Archive to Personal Assistant. We have what might be a useful desktop utility that replaces your digital diary, helps keep track of information locally on your computer, and helps manage your interactions with the internet.

Add a few decades of full-life archive and some pattern matching capability that maps patterns in our behaviour – how our behaviour and metabolism react to particular social and physical contexts – and we are looking at something beyond a simple desktop utility. We have something that in a very real sense is capturing ‘us’ – in the sense that someone who knows

you well might have difficulty telling whether they were interacting with your PA or directly with you – if they don't simply ask.

The question of PA fidelity cuts deep into our idea of who we are and what makes 'us' – our spirit. It is too big a topic to broach in detail here but one that I've addressed elsewhere. I'll make a few brief comments because it is central to how we view a PA and whether some kind of digital afterlife is meaningful.

I read quite widely in philosophical and theological writings on the subject of 'the spirit' when I was young and found little in either that could pass as clear and meaningful. It was, I think, as I waded through the spiritual abyss of Sartre's *Being and Nothingness* that I came to the conclusion that the underlying purpose of these writings was to fill a blank in our verbal sense of existence by creating forms of words that seemed to be tackling the issue but contained little, if any, substance – in a word: sophistry – well meaning and sometimes desperately so, but sophistry none the less – philosophical pillow stuffing.

I've settled for the simple common-sense view drawn largely from the way we use the word 'spirit' in everyday language. We can realistically imagine a business meeting opening with a comment from the chair: '*John can't be here today but I hope he'll be with us in spirit so the meeting can be as productive as usual.*', then someone later asking '*What would John say about that proposal?*' We might almost be at a seance. Or more generally, we can consider someone saying '*That's the spirit!*' when a friend snaps out of a miserable mood.

At a personal level, as a beautifully intricate colony of amoebal clones, we start life with the genetic inheritance of our

initial neural connections – instincts – which may shift as epigenetic factors change through our lives. Then we proceed to absorb an enormous amount of information from our senses, including the physical behaviour and mannerisms of others that we view from our own unique perspective – building models of their behaviour – generalising and categorising them. We do this by forming highly complex webs of associations – associations built from patterns detected in our sensory inputs by neural networks doing what their primary emergent behaviour leads them to do – detect patterns.

We combine instinct and association-mapping as intuition – constantly reverberating patterns of association of which we are only consciously aware of the most dominant. To cope with the complexity of the world around us we depress the clamour of diverse associations and filter out all but the most significant linear temporal stream. We can see this dominance in our visual or auditory fields where we focus on the most significant elements but the others are still in the background. Even when we sleep, with sensory inputs heavily depressed, we can still react to very particular sounds such as a baby's cry.

Perhaps uniquely as humans we then use a silent verbalisation to filter this dominant stream of associations with verbal logic. This logic can vary from a vague representation of the highlights of our continuous and diverse intuitive activity, through to the strictly defined logic that is possible in special cases.

Our internal ‘us’ develops from our interactions with the physical world and social settings where, in counterpoint, others develop views or mental models of us based on their interactions and personal perceptions of us as we contribute to the lives of those around us and to society in general. The

internal and external facets of ‘us’ are, of course, tightly interwoven. When we die they are wrenched apart and the first ‘us’ goes but the second continues – how strongly, and how far into the future, depending on how we have lived our lives and the impact (positive or negative) that we have had on others. This overview is not only a rough, modernised interpretation of the Buddhist worldview but can, I think, also be found buried in metaphor, anachronisms, and ritual in the worlds major religions. If our impact has been positive people will cling to our memory – sometimes using icons or ritual to enhance the process. If negative, people will attack, reject, or just ignore everything we stood for. We will enter heaven or hell and, perhaps in-between, a purgatory of indifference and ambivalence.

A full-life PA with metabolic monitors will pick up the instinctive, metabolic and emotional basis of our reactions or inclinations as it shares our life experiences, and learn to emulate our personal self. To do this it doesn't need to precisely map the neural connections of our individual brains but just capture the general trends of our mental behaviour as they evolve through our life.

Can it be conscious? Again, I think the discussion of consciousness has been obfuscated by sophistry. A simple, or less other-worldly answer – rejecting the modernistic mysticism of Cartesian Dualism – may be emerging from our growing understanding of brain function through the interplay between instantaneous perception and awareness, short-term memory, long-term memory, and the process of selecting a single narrative stream of awareness. This is a view that doesn't claim consciousness as something unique to humans – just

another little extension of the Copernican revolution shifting us from a unique place in Life's creations. Even the presumed uniqueness of our verbal reasoning can be questioned. Can we be sure that a dog that has learned dozens of human nouns and a few verbs does not, as we do, use them to help structure its intuitive reasoning? Can we extend that to all animals that use language?

Back in the practical world, in the lead-up to the use of PAs there is some useful functionality that is technically possible for public recording with current systems – standard privacy protocols. These would include the ability to mask out and ignore all but specified voices – either in the recording stage or recall. Similar techniques can also be applied with visual systems – for example, automatically masking all, or selected, images of people. Protocols can include standardised signals – both visual and radio frequency – to indicate when a camera or mike is operating and what privacy mode it's in. Your PA should be able to detect that you are about to move into someone's recording field and negotiate an appropriate mode.

The issues of privacy and identity can be resolved if you have a trusted entity that not only recognises your voice but can physically sense that it is really you speaking and not a recording.

So far I've described what I think is possible now or in the near future. What I've said stems from three decades of work in the area including the construction of several Natural Language Inference Engines and messing about modelling neural networks. Moving into a more speculative mode: what might the near to distant future consequences of this technology be?

A trusted PA could answer questions such as '*Where were you at the time of the crime?*' or just '*Did you commit this crime?*' or '*Are you having an affair?*'. It's not hard to imagine that this technology will radically change the way we live and interact, but how we use it can be under our control moment-by-moment.

The PA has the power to decentralise the internet by distributing more functionality and power to the nodes of the net (us), countering the centralising tendencies of monopolistic corporations and totalitarianism in governments. There are clear-cut reasons from systems theory why networks are extremely flexible and reliable and why centralisation negates those advantages and worse. *Peer-to-peer is the natural network way*.

The spectre of an all-conquering super intelligence has haunted us for many years – now in the form of the 'singularity' when machine intelligence exceeds ours and we become captive dependants to some centralised monster that's machinations we no longer understand. We can go that way if we want, but we do have choices.

As was hinted in the initial definition of PA, physical machines (bots) can be constrained to being just a physical manifestation of a Personal Avatar. Our perception of what we are dealing with and who is responsible for it is important. I've said many times over the years, '*Don't curse the machine, curse the programmer*' or more generally, the designer. Whether we are dealing with computers, cars, or any other human tool we are interacting with the people who made them.

There is a line we cross when we make automata truly autonomous (and a further one when we let them replicate). After decades considering the problems associated with

automatous systems I can see no way other than the PA for giving us a workable solution to the problems of privacy and accountability.

With every PA you get an entry ticket into the PA afterlife. Your PA can ‘live’ on after you. If another person or Trust is prepared to take responsibility for it after your death, it can continue to operate indefinitely. Your descendants and others can draw on your life experience and any wisdom you might have gleaned from it.

But there's more! Many of us like to think that humans will eventually explore the galaxy, but lugging our bulky and fragile organic forms across those vast distances and times is unlikely to be the first approach and may never happen. Forget wormholes. Faster than light travel is not even science fiction but pure fantasy. Even a flypast of our nearest star systems will take centuries. We will have to use robots but the task of pre-programming these for the many unknown problems they will encounter visiting alien planets is overwhelming. How better than to have the independent abilities of thousands of mature PAs with their combined practical knowledge and creativity? Each PA commanding a fleet of tiny craft designed to provide longevity and flexibility through massive redundancy.

If some future archeologist specialising in the silicon era happens upon a collection of crude twenty first century storage devices, pieces together your archive then activates it, they might decide to include you in their PA on an exploratory trip around the galaxy. You may not have much relevant technical knowledge to contribute but you might be able to resolve a dispute over the meaning of some song lyrics from your era

and earn enough hydrogen to be independently active at the swarm's next port of call.