AI Magazine Volume 6 Number 4 (1985) (© AAAI)

Reloading a Human Memory: A New Ethical Question for Artificial Intelligence Technology

Kenneth Mark Colby M.D.

Neuropsychiatric Institute, University of California School of Medicine at Los Angeles, 760 Westwood Plaza, Los Angeles, California 90024

Once upon a time—this phrase usually signals that what follows is a fable—there was an old professor who worked in the field of artificial intelligence applications in psychiatry and neurology. One day a man, who had lost much of his long-term episodic memory, consulted the professor to ask him if there was any way he could help him regain the lost memories.

During the previous year, this amnestic man had suffered a stroke in his right cerebral hemisphere. Being right-handed and left-hemisphere specialized for language, he was still able to speak, to read and write, and to understand what was said to him. Besides the usual difficulty in recalling proper names, his main problem involved large gaps in his memory for events that he participated in before the stroke, although he could remember events that occurred after the stroke. For example, many years before his stroke, he had received a high award for an exceptional achievement. He could not, however, remember the award ceremony nor even what it was for.

But why should he care so much? It might be a great relief to us not to remember some things. He cared because in self-referent caring about oneself, memory of past events is one of the most important and heavily weighted properties used for a sense of personal identity by fixing a previous self as one's closest continuer (Nozik, 1981). Lacking the memory-based psychological continuity and connectedness essential for personal identity (Parfit, 1984) and unable to participate fully in human conversations that make up the bulk of social life, he felt empty and suffered from feeling incomplete as a person.

The old professor deliberated for a while regarding this remarkable problem and then proposed an attempt at a solution using a simple computer-based method for trying to reload a human memory. Since there were numerous documents (diaries, letters, newspaper articles) regarding events in the man's life, and since his wife and friends were available as sources of information, the plan was to enter "stories" of pre-stroke episodes into a personal computer. Using an ordinary text-editing algorithm and a variety of changeable key words, the man could call up stories on his personal computer, read them aloud, and thus attempt to store them in those parts of his brain in which the accessing mechanisms were still intact. (From an AI standpoint, the text-editing method is trivial, but this is not an article about method; it is about ethics.) The hope was that not only would the man now have some memory to think about and talk about but, more importantly, this repeated daily practice at his own pace, with no one looking over his shoulder, might help open up new access paths to his own memory of these events, filling them in and modifying them over time.

The man talked it over with his wife and agreed to try out the plan. Each week the professor received written materials regarding events in the man's life. These "stories" were entered into files retrievable by key word mnemonics and the diskettes were sent back to the man so he could practice on them at home. If he had repeated trouble remembering a story, its accessing key-words and/or contents were slightly changed (a situation or an event can be described in an.infinite number of ways) in an attempt to stimulate multiple access pointers in his own memory.

The man realized that he was the likely murderer . . .

Now, to take up the fable, the old professor had a double motive for participating in this humanitarian effort. It seems that a few years ago he had done away with an annoying rival. The murder was never solved but the police were still on the trail. So the professor saw a chance to fit the details of the murder into the amnestic man's past life so well that when a newspaper account of the old unsolved murder appeared, the man realized that he was the likely murderer because of the time, place, and circumstances involved. He turned himself in, was convicted on the circumstantial evidence, and sentenced to prison.

I began all this with "once upon a time" to signal a fable. But the story is quite true, except for the last paragraph. The professor did not murder anyone. In fact, I am the professor. For reasons of confidentiality, the amnestic man must remain anonymous. It is, of course, quite incredible that anyone would accept such an extreme story about himself and be convicted of murder in this way. I added this twist to the story to dramatize a new ethical question for artificial intelligence technology, a field that is contributing to the development of novel methods for accomplishing things not previously achievable.

The amnestic man might have simply used a notebook, you say. The notebook would have to be the size of a building and it would take forever to find anything in it. Also one cannot easily edit stories in a notebook, continuously modifying them to provide changing pointers to access paths of memory. We tend to underestimate how large human memories are and we know little about their compartmentalization and organization for optimal accessing.

We tend to underestimate how large human memories are ...

Notice that it was not the man's own constructive and reconstructive memory that was being reloaded but other people's constructions and paraphrasings of what happened. Instead of having direct knowledge of his past, he acquires indirect knowledge, *i.e.*, knowledge *that*, with the hope that it will connect up with his direct knowledge. He may or may not accept the stories as true of him. To accept the stories as true he must trust someone that this acceptance is potentially beneficial rather than harmful. If we now have the ability to reload human memories in this way, who is to be trusted not to instill "false" memories? Who has what rights here, where does the responsibility lie? What are the ethics of this risk-benefit situation? How does one weigh the utilities of memory against the potential disutilities of false memories?

As Polanyi put it, "No intelligence can operate outside a fiduciary network" (Polanyi, 1958). Artificial intelligence involves a research community embedded in a social community having values of approval and disapproval regarding what we do. We ourselves have scientific as well as personal beliefs, goals, and values. There is a fiduciary component to scientific as well as to helping professions. The public has confidence that scientists say what they believe to be true, and do what they believe to be correct, to the best of their knowledge based on publicly scrutable evidence. The public must be able to trust scientists and technologists to exercise responsibility.

With the great amount of attention now being paid by the media to artificial intelligence, it would be naive, shortsighted, and even self-deceptive to think that there will not be public interest in scrutinizing, monitoring, regulating, and even constraining our efforts. What we do can affect people's lives as they understand them. People are going to ask not only what we are doing but also whether it should be done. Some might feel we are meddling in areas best left alone. We should be prepared to participate in open discussion and debate on such ethical issues.

Ethical monitoring should come from colleague control. It can only be hoped that we can avoid federal watchdog agencies or any other bureaucratic structure interested in power manipulations and authoritarian pressure. A lamentable example is the recent case of recombinant DNA in which apocalyptic arguments about disaster were based entirely on ignorance (Holton and Morrison, 1979). In combining rules and cases, it is difficult to decide which rule to apply to specific cases having multiple variables and great variability of circumstances (Abernethy, 1980). Each case must be evaluated according to the informed needs, interests, and abilities of the participants. It is they who bear the ultimate responsibility of personal judgement and there is no way of taking that burden away from them, even by a committee.

References

Abernethy, V (Ed), (1980) *Frontiers in medical ethics: applications in a medical setting* Cambridge: Ballinger

Holton, G., & Morrison, R S (Eds), (1979) *Limits of scientific inquiry* New York: W W Norton.

Nozick, R (1981) *Philosophical explanations* Cambridge: Harvard University Press

Parfit, D (1984) *Reasons and persons* Oxford: Clarendon Press

Polanyi, M. (1958) *Personal knowledge* Chicago: University of Chicago Press