The Knowledge Engineering Review, Vol. 12:3, 1997, 229–230

Guest Editorial

DAVE STUART ROBERTSON

Knowledge based systems are used in applications where an incorrect decision could put human life in jeopardy. A quick trawl through the World Wide Web is sufficient, these days, to locate such applications in design, analysis and testing; protection advice; operator decision support; signal monitoring; embedded systems and others. Depending on the type of system, these either give information which is not guaranteed to be correct (in many operator support applications) or which is imprecise (for example in fuzzy logic controllers).

This is not always a bad thing. An imprecise answer may be sufficient for making appropriate decisions. Occasional inaccuracies may be acceptable in systems which support rather than supplant operators. Whether or not our systems create a safety problem depends on the context in which the expert system is deployed, which should constrain the architecture used to build the system and the ways that we argue for its safety. This issue of *The Knowledge Engineering Review* views this issue from four perspectives:

- *Proof versus empirical argument:* although traditional safety engineering aspires to precise understanding of the potential behaviours of deployed systems, there are sometimes unpredictable hazardous situations where this is impractical. Neural computing methods offer new choices of architectures for such problems, but bring with them the need for different ways of arguing for the safety of systems, based on statistical arguments. In particular, we would be concerned to estimate the robustness of the system to new inputs on which it wasn't trained. Sharkey & Sharkey discuss what these empirical arguments might be like.
- *Communication:* when producing specifications of systems or articulating formal arguments for their safety, a formal presentation (even if correct) may be unconvincing unless it is presented in a form which is accessible to human inspection. Many different people, from differing engineering cultures, may have a legitimate interest in the design of a system so a single way of discussing the design is unlikely to suit everyone. Gurr's paper looks at the ways in which different forms of communication have been used to communicate designs and the extent to which it is possible to relate successful forms of communication to formal theories.
- *Integrating with established methods:* it is tempting sometimes to think of knowledge based systems as ways of revolutionising work practices. In yielding to this temptation we may become blind to the beneficial features of existing methods, and the engineering cultures which have grown around them. Price *et al.* give an example of how this problem may be avoided. They describe how qualitative modelling techniques can be integrated with an established failure modes and effects analysis, enhancing safety analysis without displacing existing good practice.
- *Reasoning about system safety:* arguments which help us to understand whether a system has been safely designed are as important as the system itself. Given that these arguments require the deployment of knowledge and expertise it is natural to consider whether some aspects of the articulation of safety arguments might be supported by automated methods. Krause *et al.* survey some of the approaches which have been used: linking formal expressions to requirements descriptions in controlled natural language; using formal requirements to constrain and endorse design; and reasoning formally about the uncertainty in safety arguments.

A common theme running through all the papers in this issue is the need to understand more fully how technical methods used in constructing knowledge based systems can be embedded within

acceptable safety engineering cultures. This is a problem for the knowledge engineering community, not because our methods lack rigour or precision (many are at least as precise and rigorous as more conventional software engineering methods) but because they are often targeted at parts of problem domains which are, themselves, imperfectly understood.

https://doi.org/10.1017/S0269888997003019 Published online by Cambridge University Press