

# Interview mit David Leake



**David Leake received his PhD in Computer Science from Yale University in 1990. Since that time he has been on the faculty of Indiana University, Bloomington, where he holds the positions of Professor in the Computer Science Department and Associate Dean for Graduate Studies in the School of Informatics, as well as being a member of the faculty of the Cognitive Science Program and Human-Computer Interaction Program. His research interests are in artificial intelligence and cognitive science, including case-based reasoning, explanation, intelligent user interfaces, and knowledge management. He is the Editor of AI Magazine and he has authored respectively co-authored over 100 publications in these areas including a book entitled "Evaluating Explanations – A content theory".**

*KI: Explanation is a rather old fashioned topic in AI. While it was quite hot in the years of expert systems, it was considered unfeasible in the years after. What was the reason for this?*

Very early, there was an insight that explanation was crucial for system acceptance. I remember that in Buchanan's and Shortliffe's early work on medical consultation systems, the ability to create explanations was a number one requirement. Initially, it seemed that simple explanation mechanisms might be sufficient, and tracing the rules used by an expert system is straightforward. Fairly soon it became clear that what makes a good explanation depends on much more than simply tracing the systems internal processing. There are all sorts of different types of explanations, which makes it necessary to develop some deep understanding of the nature of explanation first and then to develop AI methods to reflect all those different types.

---

## *Explanation became a fundamental AI problem in its own right*

---

Furthermore, explanations are very task sensitive and they depend on the particular interaction with the user. All these things changed perceptions and revealed so many rich issues that explanation became a fundamental AI problem in its own right. Many issues must be addressed to achieve explanation systems that can tackle the full range of explanation problems.

*KI: What aspects are important to understand the nature of explanation?*

The key thing that we have come to understand is that explanation must be very context sensitive. It also must be very goal oriented because different users will have different information needs and the explanation process really has to address those needs. Explanation also depends on the explanations' purpose. For example, it makes a difference whether we want the explanations to increase a user's trust in a system decision versus if we want educate the user about how the system works or the subtleties of the domain. The user's needs may be quite divergent which also requires different types of explanations.

Early explanation systems treated the problem of explanation as a one way transmission of information from the system to the user. This was a useful assumption to keep systems simple. However, in the longer term it will be much more worthwhile to have an interactive and collaborative process that lets the user personalize the process on the fly.

*KI: Which expectations concerning explanation have been unrealistic in the old days?*

One of the expectations was simply that a low level trace fairly close to the system implementation would be sufficient for explanation. If you have a simple task and a simple domain, that may be adequate. Certainly that is very useful for either system developers or for those who already have a fair amount

of expertise with the system. But if you move into more complex domains, presenting all the low level rules may not be very informative to a user.

Nor should the aim of explanation necessarily be to explain the system. I remember that there was an early debate on how faithful explanation should be to the system processing. The question was: "should the explanation be very closely tied to what the system did or should it be decoupled, to present information in a way that might be easier to understand but farther from what the system actually did?" I think it's clear now that what matters is addressing user needs.

---

## *There was a lack of cross-fertilization which probably slowed progress*

---

*KI: Progress in explanation systems looks rather slowly, even in the years after the expert system hype. What were the reasons for this?*

Later on, there was a lot of excitement in AI for black box methods like neural networks. These methods are hard even for people to explain, but they perform well, so there may have been an implicit assumption that if the performance was good enough, that was all that mattered. However, users are understandably very reluctant to make certain very critical decisions if they have no idea why the system is actually making recommendations. One of the challenges for explanation now is how to generate better explanations with those systems.