The Use of Legal Knowledge-Based Systems in Public Administration: What Can Go Wrong?

Hugo de Bruin[1]   Henry Prakken[2]   Jörgen Svensson[3]

[1] *Cognitive Artificial Intelligence Programme,*
Utrecht University, The Netherlands
[2] *Institute of Information and Computing Sciences,*
Utrecht University, The Netherlands
[3] *Faculty of Public Administration, Sociology Department*
University of Twente, Enschede, The Netherlands

Abstract. In recent years, practical applications of legal knowledge-based systems have become increasingly common. This raises the issue of their functioning in practice and their actual influence on the quality of decisions. In this paper we investigate to what extent incorrect decisions may be caused by factors that cannot be attributed to flaws in the programme's knowledge base or reasoning. Based on a literature study, five possible causes are identified that pertain to the interaction between a programme and its user. Then it is illustrated how this list of causes may be used to investigate risk factors in practical applications, with a small case study on the use of a knowledge-based system in processing claims for general income support in a Dutch municipality.

1 Introduction

In recent years, practical applications of legal knowledge-based systems (LKBS) have become increasingly common in, for instance, public administration ([6, 12]), prosecution assistance ([14]), legal advice ([13, 9]) and e-government (see e.g. www.elaws.com). This raises the issue of the quality of their actual functioning in practice. It seems safe to assume that LKBS applied in practice do not work flawlessly. In a recent study by [5] it was found that the Dutch MRE system for social welfare benefit, although generally functioning satisfactorily, still produced some incorrect results. Of course, this is not a reason not to use the programme: what counts is whether decisions taken with the help of an LKBS are less often mistaken than decisions taken by humans alone. It is known from several studies (e.g. [1]) that human officers processing social welfare claims make a considerable number of mistakes, often with financial consequences for the clients. Both laboratory studies (e.g. [11]) and practical experiences (e.g. [6]) indicate that with an LKBS the quality of decisions can be considerably improved.

For this reason, it becomes even more important to investigate possible problems in the use of LKBS. In order to further increase the quality of the decisions made with their help, a clearer picture is needed of possible sources of errors. One such possible source is, of course, shortcomings of the LKBS itself, either in its technical or in its legal aspects. Indeed, [5] found that some of MRE's incorrect results were due to flaws in its legal knowledge base. However, other errors could not be attributed to the LKBS itself. In this paper we discuss another possible source of errors, viz. sub-optimal interaction between an LKBS and its user.

Based on a study of the literature on legal knowledge-based systems and human-computer interaction, our aim is to provide a classification of possible ways in which sub-optimal interaction between an LKBS and its user may lead to incorrect decisions. We also discuss some practical difficulties that evaluation studies will face when defining the notion of an incorrect outcome. One possible use of our classification of problem causes is the development of a checklist that can be used to identify risk factors in a concrete application. We illustrate this with a case study where the use of an LKBS at a municipal social welfare department was investigated.

The focus of our research is limited in two respects. Firstly, we focus on LKBS that support determinative processes in public administration. Determinative processes (cf. [6]) are processes where a 'street level bureaucrat' decides about the rights and obligations of a particular citizen on the basis of legislation (cf. [8]). Typical examples are social welfare legislation, building permits, and tax assessment. Secondly, as this is only a first investigation meant to do some groundwork for further studies, our aim is to obtain a first insight rather than an exact quantification of the problems and their causes.

2   What is an incorrect decision?

In this section we discuss where in the process of using an LKBS errors can occur, and we discuss some practical problems with defining and identifying incorrect decisions.

2.1   Types of legal errors

From a legal point of view, two types of errors can be distinguished. Decisions can be materially incorrect, i.e., wrong about the rights, obligations and powers of legal subjects, and they can be formally incorrect, i.e., they may contain errors in their wording (e.g. a decision does not contain all the grounds on which it is based) or in the way they were taken (e.g. a deadline was violated).

2.2   Where can errors be made?

When a case processed with an LKBS is incorrectly decided, the problem may arise in several ways (see also [11]). Sometimes it happens that on the basis of correct input, the programme produces incorrect output. In such cases, the programme's knowledge base and/or reasoning mechanism is flawed. However, even if the programme's knowledge and reasoning is adequate, decisions taken with its help may be incorrect. This may especially occur in two types of situations. Firstly, the input can be incorrect, that is, the data as provided by the client or some other source may be incorrectly entered into the system by the consultant. Secondly, it may happen that the programme's correct output is used incorrectly in making the final administrative decision. As already said in the introduction, in this study we only focus on cases of the latter two types, assuming that the internal functioning of the LKBS is adequate. However, drawing the borderline between the various cases is not easy, especially since most (if not all) LKBS for public administration provide only partial support to their users, so that it becomes hard to tell whether a mistake must be attributed to the system, to the user or to both.

2.3   Partial support

For several reasons the support offered by an LKBS will usually be partial. Sometimes an LKBS only produces the material aspects of a decision, and leaves it to the consultant to extend it to a full text that meets the legal demands on wording and form. Another source of partial support is that administrative laws generally leave room for judgment by the civil servants. For instance, concepts used in regulations may be vague or open-textured, so that their application to concrete cases requires judgment. Although an LKBS could provide partial support in the form of, for instance, natural language explanations, lists of relevant factors, examples, or precedents (cf. [6]), it has to leave the final judgment to the consultant. Regulations may also deliberately leave certain decisions to the discretion of the decision-maker, and administrative guidelines for exercising this discretion may not be included in the knowledge base. [5] discuss the following example where a discretionary decision is incompletely supported by an LKBS. The Dutch general assistance act gives the municipalities the power to reduce or cancel a benefit if an applicant fails to meet certain obligations. However, the act leaves it to the municipalities' discretion to apply this sanction; all it does is specifying the circumstances that must be taken into account in the decision. The MRE-Abw system (see further Section 5 below) gives incomplete support of this decision-making task: it produces a draft text for the decision but it does not support the process of taking the relevant circumstances into account. A final reason why support by an LKBS will never be complete is that specific regulations may be set aside by higher general statutory norms or even by unwritten legal principles. For instance, in the Netherlands, all public administration decisions are subject to the so-called general principles of proper administration, codified in the General Administration Act. Most of such higher legal sources are even more vague and open-textured than the 'primary' regulations, and they are often not included in the programme's knowledge base. All these features of legal systems make that an LKBS simply cannot support every task its users may be confronted with.

This observation has some practical consequences for evaluation studies of LKBS, since it implies that the borderline between incomplete and incorrect support is vague. Consider the case where a programme does not contain the general principles of proper administration (GPPA). If it qualifies every output with 'subject to the GPPA', then it may be clear to the user that s/he must still assess the output in light of these principles, so in this case the programme provides correct (albeit incomplete) support. However, what if the programme only displays one general message every time it is started? Or if it does not mention the GPPA at all but they were pointed out to the consultant during his training? Or if the managers in the social welfare department are aware of the incompleteness but do not inform the end-users? Clearly, there is a grey area between incomplete and incorrect advice by an LKBS, so that there is a danger that different evaluation studies give different meanings to these terms.

2.4   Overruling regulations on social grounds

Evaluations of LKBS also face another problem. In general, there is no 'gold standard' for what is a correct legal decision: lawyers can reasonably disagree about many things, so that it often depends on the expert consulted whether a decision is regarded as correct or incorrect. Similar problems are well known in the evaluation of medical knowledge-based systems; see e.g. [15]. When the LKBS operates in public administration, this problem may arise in a specific form. Sometimes consultants deviate from the regulations since they feel that their outcome in a specific case is unfair or socially unacceptable. This is a well-known phenomenon in street-level bureaucracies (cf. [8]), caused by the fact that the relevant legislation

is often complex, vague or even inconsistent, so that their application to individual cases may have unforeseen consequences. [7] discusses several interesting cases where she found that consultants wished to deviate from a regulation. For example, an abused woman in hiding for her husband received a temporary benefit although she was unable to identify herself with an official document. And in a controlled laboratory evaluation of the TESSEC system, [11] found three cases where consultants deliberately modified the input in order to obtain an outcome that in their view was more fair for the client. Their underlying motivation was that the relevant legislation was poorly designed so that it treated several categories of clients unfairly. Now the question arises whether in such cases the decision must be qualified as socially desirable but legally incorrect, or as legally correct since socially desirable. Clearly, this depends on several factors, including one's philosophy on the role of social values as sources of law. Again there is a danger that different evaluation studies give different answers to this question.

3   Possible causes of errors

Keeping the observations of the previous section in mind, we now discuss some possible causes of incorrect decisions produced with the help of an LKBS that in itself functions correctly. Based on an analysis of the literature in two fields, human-computer interaction and legal knowledge-based systems, we present a classification of five possible types of causes.

1. *Insufficient domain knowledge to verify and adjust LKBS outcomes.*

   Above we have seen that an LKBS usually provides only partial support, so that the consultant must complement the system by using his or her own legal skill and knowledge. This leads to the first source of possible errors, namely, that the consultant does not have the legal knowledge needed to take the legal decisions that the LKBS leaves open. This includes lack of knowledge of the relevant legislation but also exercising legal judgments in a way that domain experts regard as flawed on the basis of, for instance, legal principles, case law or standard policy. Several cases of the latter type (ignoring standard policy) were reported by [11].

2. *Insufficient awareness of the extent to which the LKBS supports the user's task.*

   The partial nature of the support offered by many LKBS also leads to another possible source of errors. As discussed above, a consultant may not be aware that s/he should verify or extend the programme's output and may therefore incorrectly rely on its completeness and correctness. Several cases of this kind were found by [5]. As discussed above, such cases must sometimes be regarded as incorrect output of the LKBS. However, if the programme cannot be blamed, then the consultant's reliance on its output may have several deeper causes. For example, the consultant may lack sufficient legal knowledge to recognise the partiality of the support. This is in fact a variant of (1). Also, the consultant may have received insufficient information about the functioning of the programme. Finally, she may have too much confidence in the LKBS as a piece of technology. Unwarranted reliance on technology has been reported several times in the literature. For instance, [3] found in a controlled laboratory study that law students attached more value to incorrect advice by an LKBS than to correct advice provided without an LKBS.

3. *Strategic behaviour.*

An incorrect outcome may also be due to so-called strategic behaviour by the consultant. Quality reviews on the use of an LKBS in an Australian government department identified two types of consultants ([6]). One type relied on the programme when appropriate and complemented it when needed, without preconceived ideas about the desired outcome. However, the second type of consultant tended to work on the basis of a decision made in their own mind, and then used the LKBS to reach that decision. Although such 'strategic behaviour' may very well lead to correct decisions, it may also lead to incorrect ones, for instance, if a consultant wants to deviate from the regulations for reasons of fairness in a way that must be regarded as legally incorrect.

4. *Usability problems.*

A fourth possible source of errors is that the user does not make optimal use of the software because of the way it is designed or because of the way its knowledge is modelled. The first aspect concerns a programme's usability. According to [10] a piece of software is usable if it complies with the following criteria:

Learnability: The programme should be easy to learn. Metaphors (desktops!) and consistent interface design are two well-known principles here.

Efficiency: The programme must be efficient in use in that the processing of a case should take less time using the programme than without.

Memorability: Users should be able to remember the working and functioning of the programme easily. Again metaphors and consistent interface design can be of help.

Few errors: The programme should work in an error-avoiding way, and if an error is made, there should be an easy way of correcting it. A programme therefore has to operate in a logical, predictive way.

Satisfaction: Finally, the user should be satisfied with the use of the software.

Based on these criteria, Nielsen proposes a method for usability engineering that can be done relatively easy and with a small budget. For present purposes the 'error avoiding' aspect is most relevant. Among other things, it means that 'mechanic', thoughtless use of an LKBS should be discouraged, for instance, by avoiding excessive use of default values.

As for the way the system's knowledge is modelled, a mismatch may occur between the user's and system developer's mental models of this knowledge; see e.g. [4]. For instance, they may differ in their interpretation of certain concepts.

5. *External pressures.*

Finally, stress-producing external factors, such as a high workload, time pressure, poor management or aggressive behaviour of clients, can cause errors, either alone or in combination with other possible causes. For instance, time pressure may lead to incorrectly accepting default values in cases where these defaults do not apply, or it may lead to consultants failing to fill in the parts the LKBS does not support.

4   Towards a checklist for practical applications

What can be done with the insights of the previous sections? From a scientific point of view it is necessary to carry out empirical field studies as to whether the possible causes of errors

indeed occur in practice. If they are found, it should be investigated to what extent they can explain errors that are found. However, in this paper we discuss a more practical use of our insights, in supporting persons responsible for practical applications of LKBS to recognise potential dangers. To this end, our list of possible causes of errors could be used to formulate a checklist with which potential risks can be identified. Such a checklist could, for example, contain the following questions.

1. *Insufficient domain knowledge to verify and adjust LKBS outcomes*

   - Are the users of the system legally qualified to make these decisions?
   - Do the users have sufficient experience with the task to be performed?
   - Are the users (still) able to make the same decisions without the system?

2. *Insufficient awareness of the extent to which the LKBS supports the user's task*

   - What knowledge is built into the LKBS?
   - To what degree is the knowledge base complete with respect to the tasks of the user?
   - Are the users sufficiently informed about the tasks the system does and does not support?
   - Is the output clear about the exact legal basis of the advice?
   - Does the LKBS prompt the user to make certain decisions himself?

3. *Strategic behaviour*

   - Do the users acknowledge the legitimacy of the regulations in question?
   - Are the users socially and/of emotionally involved in the decision? (for instance, are they in personal contact with the citizens?).

4. *Usability problems*

   To evaluate usability of the system a standard method from the HCI literature could be used, such as the quick and cheap method of [10]. Some other relevant questions are:

   - Did the users receive sufficient training in the use of the software?
   - To what extent does the software encourage mechanic, thoughtless processing of a case? For instance, to what extent are default values used?
   - Is the output of the LKBS clear about the input that was used, so that input-errors may be recognised afterwards?
   - Can input errors be corrected easily?

5. *External pressures*

   - What is the workload of the users, do they have time to make their own judgments or is the use of the LKBS a repetitive task?

5 A case study

We now illustrate our remarks on the practical use of our observations with a case study (described in more detail in [2]). We performed an interview-based survey within a Dutch municipality that uses a knowledge-based system called MRE-Abw, which supports the application of the 'General Assistance Act' (Algemene Bijstandswet, Abw). The Abw regulates the provision of income support by municipalities for citizens who lack sufficient means to support themselves. It is characterised by a considerable number of open-textured terms and substantial room for discretion, although some of the room for judgement and discretion is constrained by local municipal guidelines. The MRE programme is used by a consultant who gathers the data in interaction with the client while using the programme. Some of the data may be imported from other programmes. The programme's output is a word processor file containing a draft decision, which can be manually edited by the consultant. Each decision is reviewed before it is finalised (although in the investigated municipality the decision maker and reviewer are the same person).

In our case study we investigated whether there are indications of possible causes of errors. Moreover, to find out whether our checklist indeed helps to identify risk factors, we also gathered information on the actual occurrence of errors in the studied municipality.

5.1 Possible causes of errors

In order to identify possible causes of errors, the software was investigated and six consultants and three staff members were each interviewed for one hour. The results of this investigation can be summarised as follows.

1. *Insufficient knowledge to verify and adjust LKBS outcomes*

   The consultants had a working experience within the department ranging from one to six years and they had all received considerable legal education (two of them at academic level). Five of them had followed additional courses on processing administrative legislation. All consultants thought they had sufficient legal knowledge to process a case with the programme. This was confirmed by the internal quality reviewer. The consultants also indicated that they thought to have sufficient knowledge of the domain to process a case without support of MRE (although only one of them had ever done this). In sum, it seems that the consultant's level of legal expertise as needed for this domain is quite high.

2. *Insufficient awareness of the extent to which the LKBS supports the user's task*

   Although the consultants said to have high confidence in the internal functioning of the programme, they said to be less confident in the quality of its output. They reported that they had been warned during their training not to fully trust the programme, and said that they complied with this advice. Five of the six consultants also reported a general awareness of the partial nature of the programme's support. Since they often used the option to adapt the programme's output with a word processor, they felt that the quality of the final decision depended more on them than on the programme. In fact, they seemed to value the system more as a checklist and an efficient way to draft a decision than for its legal expertise. Finally, all of them believed that they generally made no mistakes when completing the programme's output.

3. *Strategic behaviour*

General income assistance is typically a kind of legislation that may provoke strategic behaviour. Moreover, in the investigated municipality the reviewer of each decision is the same person as the one who drafts the decision, who has been in close personal contact with the client. This is known to be another factor that may provoke strategic behaviour ([8]). However, in the interviews we found no clear indications of this type of behaviour. Although four of the six consultants indicated that when using the system they have an idea in mind of what should be the outcome, they also said that this is due to their experience with the functioning of system. Three of them indicated that if the system's outcome was unexpected, this was almost always due to an input error. None of the respondents reported cases where they deviated from the system's output for social or fairness reasons.

4. *Usability problems*

To identify possible usability problems, the programme was assessed on the five criteria of [10]. The conclusion was that the programme scored well on all of them, although one risk factor was identified: the frequent use of default values combined with the way of navigating through the screens might give rise to a rather quick and mechanical use of the system. Although most consultants had hardly received any specific training in how to use the programme, they already worked quite some time with the software and felt that they knew it well enough. Moreover, they generally found the programme pleasant to use. We did not investigate possible mismatches between the users' and system developers' mental models of the programme's knowledge.

5. *External pressures*

The consultants felt that they have sufficient time to process a case, and they found the work atmosphere generally pleasant.

To conclude, on the one hand we found that the persons interviewed were generally rather confident about their use of the system and tended to downplay possible causes of errors. On the other hand, our checklist still revealed three risk factors: the incomplete nature of the programme's support (even though the interviewees did not regard this as a serious problem), the possibility of careless use of the system, and the possibility of strategic behaviour.

5.2   Occurrence of related errors

We next gathered information on the actual occurrence of errors in the investigated municipality. Two sources of information were available: within the municipality an annual internal quality review is carried out on the basis of a random selection of cases, and the municipality had been the subject of a field study of [5]. One of the interviewed staff members was an internal quality reviewer, who reported that in 2001 a random selection of 92 cases contained 21 cases with errors. However, according to the reviewer, none of these errors had financial consequences. Furthermore, several errors were administrative rather than legal, such as violations of internal guidelines on how to sign a decision. In addition, [5] had found that in cases where the LKBS offered incomplete support a considerable number of mistakes were made. Some of them were errors in the content of the decision, but it is hard to determine whether they had financial consequences, since they all consisted of not taking an obligatory set of circumstances into account in an otherwise discretionary decision (the example discussed above in Section 2).

In sum, although the consultants gave the impression of an unproblematic application of the system, the available evidence suggests that a still a significant number of errors is made. Of these errors, those reported by [5] are clearly related to one of the risks identified in our case study, viz. the problem of incomplete support, as well as to the fact that the consultants seem insufficiently informed about the partial nature of the support. Furthermore, the quality reviewer reported errors of carelessness of the consultants, which may well be connected with one of the other risks identified in our case study, viz. that of careless processing of cases.

5.3 Discussion

With respect to the investigated municipality it can be concluded that some possible causes of errors are present. The most serious risk seems to be that the consultants are insufficiently aware of the specific tasks that are left to them. Since their level of legal expertise is high and they all expressed a critical attitude towards the system, we hypothesise that this unawareness may be due to the fact that the programme is not explicit enough about the partial nature of its support, and/or to insufficient training of the consultants in using the program. Two other risk factor identified are that the system's interface may give rise to careless processing of cases, and that the fact that the drafter and reviewer of each decision are the same person may give rise to strategic behaviour. However, the latter two causes should be further investigated.

It is hard to draw more general conclusions from this study, since it is limited in several respects. Only a small number of people were interviewed, all in the same municipality and all working with the same programme, in the same field of application. Moreover, our survey has mostly provided subjective data on how the respondents perceive and interpret their use of the programme, which may very well differ from how they actually use it.

With this in mind, the following can be said. This case study partly confirms earlier findings by [7], who carried out an interview-based survey at several social welfare departments, investigating the extent to which 'hard cases' are dealt with as 'easy cases' when processed with MRE. As found both by her and the present study, consultants using MRE especially value its use as a checklist and tool for drafting a decision, and generally do not feel seriously restricted in their exercise of judgment and discretion.

The case study also gives some support to Groothuis & Svensson's conclusion that the partial nature of the support of an LKBS is an important source of incorrect decisions. Where Groothuis & Svensson found that in cases with partial support several mistakes were made, our survey, which found no strong presence of other possible causes of errors, indicates that this may be due to insufficient awareness of the partial nature of the programme's support.

6 Conclusion

In this paper we have investigated what can go wrong with decisions taken with an LKBS that in itself functions correctly. Based on an analysis of the HCI and LKBS literature, we have identified five possible causes of incorrect decisions. We have also suggested and illustrated a possible practical use of our findings, viz. designing a checklist with which a practical application can be checked on risk factors. An important result of this paper is a call to researchers, LKBS designers and user organisations to pay more attention to the problem of partial support. This problem should be addressed in the design of a system as well as in the implementation and maintenance of such a system in an organisation.

References

[1] Bien, E.A. & Roeders-Veening, C., *De uitvoering van de Bijstandswet: een vergelijkend administratief-organisatorisch onderzoek bij zes gemeenten*, Rotterdam: Stichting Moret Fonds, 1995 (in Dutch)

[2] Bruin, H.J.A.T. de, Tussen Consulent en Kennissysteem. Een Onderzoek naar de Gebruiksvriendelijkheid van Juridische Kennissystemen, Msc Thesis Cognitive Artificial Intelligence Programme, Utrecht University, 2002 (in Dutch). Available at http://www.cs.uu.nl/groups/IS/studprojects/debruin.pdf.

[3] Dijkstra, J.J., User interaction with legal knowledge-based systems, in: Beuker, Leenes & Winkels (eds.), *Legal Knowledge and Information Systems. JURIX-2000: The Thirteenth Annual Conference*, Amsterdam: IOS Press, 2000, pp. 11-21

[4] Engers, T.M. van, *Knowledge Management. The Role of Mental Models in Business Systems Design*, Doctoral Dissertation Free University Amsterdam, 2001.

[5] Groothuis, M.M. & Svensson, J.S., Expert system support and juridical quality, in: Beuker, Leenes & Winkels (eds.), *Legal Knowledge and Information Systems. JURIX-2000: The Thirteenth Annual Conference*, Amsterdam: IOS Press, 2000, pp. 1-10

[6] Johnson, P., JURIX-2000 Tutorial notes on Legal knowledge-based systems in administrative practice and electronic service delivery. (e-government), 2000.

[7] Linden-Smith, T. van der, *Een Duidelijk Geval: Geautomatiseerde Afhandeling*, Den Haag: Sdu uitgevers, 2001 (in Dutch)

[8] Lipsky, M., *Street Level Bureaucracy: Dilemmas of the Individual in Public Services*. Lexington, Mass., Heath, 1980.

[9] Macrossan, N., & S. Hart, Interactive Incorporations Website: http://incorporator.com.au *Proceedings of the ICAIL-2001 Workshop on Legal Knowledge Systems in Action: Practical AI in Today's Law Office*, pp. 54-58.

[10] Nielsen, J., *Usability Engineering*, London: Academic Press, 1993

[11] Nieuwenhuis, M.A., *TESSEC: een Expertsysteem voor de Algemene Bijstandswet*, Deventer: Kluwer, 1989 (in Dutch)

[12] Svensson, J.S., The use of legal expert systems in administrative decision making, in: A. Grönlund (ed.), *Electronic Government: Design, Applications and Management*, London etc.: Idea group publishing, 2002

[13] Weusten, M.C.M., *De Bouw van Juridische Kennissystemen, KRT: Methodologie en Gereedschap*, Deventer: Kluwer, 1999 (in Dutch)

[14] Woodin, D.E., Design and implementation of GungaWeb: an application of classical expert system technology to the production of Web-based commercial systems. *Proceedings of the Eighth International Conference on Artificial Intelligence and Law (ICAIL-2001)*. New York: ACM Press, 2001, pp. 104-108.

[15] Wyatt J. & Spiegelhalter D. Evaluating medical expert systems: what to test and how? *Medical Informatics* 15:3 (1990), 205-217.