AI Magazine Volume 9 Number 2 (1988) (© AAAI)

What AI Practitioners Should Know about the Law

Part Two

Steven J. Frank

This is Part 2 of a two-part article and discusses issues of tort liability and the use of computers in the courtroom. [The legal dimensions of topics covered in this part are given comprehensive attention by the author in Tort Adjudication and the Emergence of Artificial Intelligence Software, 21 Suffolk University Law Review *623 (1987)].*
Part 1 of this article, which appeared in the Spring 1988 issue of AI Magazine, *discussed steps that developers of AI systems can take to protect their efforts, and the attendant legal ambiguities that must eventually be addressed in order to clarify the scope of such protection. Part 2 explores the prospect of AI systems as subjects of litigation.*

Although recent computer-related disputes have involved a variety of legal issues—breach of contract specifications, misrepresentation of performance capability, and outright fraud, to name a few—it is the threat of tort liability that continues to strike the greatest fear in the heart of mighty industrialist and struggling start-up alike.

Once inside the courtroom, what role can the computer assume in its own defense or in the service of some other litigant? The law of evidence, developed to govern the testimony of human witnesses, must continually evolve to accommodate new, nonhuman sources of information. At the same time, developers of intelligent computer systems will have to achieve a great deal of communicative power before their machines will be given a day in court. These two topics—tort liability, and the computer as witness—are the focus of Part 2.

Tort Liability

The system of tort compensation exists to provide victims of injury—physical, economic, or emotional—with the means to seek redress in the form of monetary damages. The purpose of the damage award is to "make the plaintiff whole" by compensating for all losses flowing from the defendant's wrongful act. Although this concept seems straightforward, damage awards have been known to vary enormously, even for identical injuries. The host of subjective factors that appear in the calculus of compensation precludes accurate forecasts in most cases, leading some defendants to financial ruin and others to surprised relief. However, injury alone does not guarantee recovery. Rules of tort liability mediate between the victim's need for recompense and the defendant's right to remain free from arbitrarily imposed obligation.

Although sensational cases involving large recoveries tend to generate the greatest alarm, the magnitude of damages in a particular case is actually far less important than the availability of any damages in similar cases. Potential tort defendants are primarily interested in their overall liability exposure, as determined by the evolving structure of case precedent. For software developers, expenditures for debugging, design safety analysis, and quality control assurance are necessarily affected by the extent of perceived vulnerability to tort actions.

More so than in any other legal field, the boundaries separating human from machine will be forced into focus by questions of tort liability. In addition to compensation, another function of tort law is to project proper standards of care in the conduct of activities that might cause harm. When the agent of injury is a tangible device or product, attention is currently directed toward three sets of possible culprits: manufacturers (Was the product designed or manufactured defectively?), sellers (Was the product sold in a defective condition?), and purchasers (Was the product used improperly?). Absent from the lineup is the injury-producing item itself, whose existence is relevant only insofar as it pertains to the conduct of human beings. As "devices" come to include electronic systems capable of judgment and behavior, they too will become objects of direct inquiry. Naturally, financial responsibility will ultimately rest with a human being or a corporation possessing a bank account; but the standards by which humans and machines are judged will begin to merge as the tasks they perform grow similar.

The question of whether liability accrues in a given instance invariably reduces to the application of two legal variables: the standard of care expected of the defendant and the requirement that a causal link exist between the defendant's substandard conduct and the plaintiff's injury. The existence of a valid tort cause of action means that someone has suffered loss. This much is beyond change. The function of the liability standard is to allocate loss among all parties involved based on considerations of fairness, efficiency, and accepted behavioral norms. The terms of the standard prescribe the level of vigilance expected of individuals who conduct a particular type of activity. Although variations exist, most formulations derive from the fault-based concept of negligence: If the defendant failed to act in the manner of a reasonably prudent person under all circumstances, the loss falls on him or her. A special relationship between plaintiff and defendant or the performance of an unusual activity can prompt a court to adjust the standard. Innkeepers, for example, have been held liable to guests for the "slightest negligence" because of the high degree of trust placed in their hands. Certain activities have been identified as so inherently dangerous or unfamiliar that no degree of care can adequately prevent mishap. For these "strict liability" activities, the loss falls on the defendant, regardless of fault, as a cost of doing business.

This choice of liability standard calls for a decision based on public policy. Although through less obvious means, the parameters of causation are ultimately shaped by similar considerations. A scientific view of deterministic causality provides only a starting point for the legal notion of causation. Judges have recognized that too many events are logically interrelated for liability to rest solely on logic. Fairness to defendants requires that a line be drawn at some level of remoteness, and the location of this line reflects a policy-oriented value judgment. Liberal tests of causation shift a greater amount of loss toward tort defendants by including more distant effects of liability-producing conduct within the range of potential recovery. These tests generally focus on the existence of a chain of events, such that causation is based on proof of an unbroken progression of occurrences; the outer limits of connectedness are typically bounded only by the reluctance to impose liability for the extremely unpredictable. Restrictive tests of causation focus directly on the defendant's ability to foresee the possible harm arising from particular actions and reverse the shift in favor of the plaintiffs.

Returning to liability issues, the rule applied to parties engaged in commercial trade depends heavily on what is being sold. Providers of services have traditionally been held to a negligence standard based on reasonably prudent practice within a given area of specialization. In contrast, sellers and manufacturers of tangible commodities face strict liability for injuries to individual consumers caused by defects in their wares. The reason for the distinction does not lie in any perceived disparity of associated danger, but rather in the fact that product sellers are viewed as economically better able to spread a loss over a large number of users through price adjustments.

Courts have had difficulty fitting computer programs into this bifurcated world of products and services. Software can be supplied in a variety of forms, some more tangible than others. Programmers might write software for mass distribution, tailor an established package to the needs of a particular user, or design from scratch a custom system for an individual client. Although no court has yet faced this issue in the precise context of tort liability, most commentators (including this writer) believe that characterization as product or service is most properly determined by the supplier's degree of involvement with the customer. Greater availability of technical assistance and support make the overall transaction appear more like a service.

Because implementation of the high-level tasks performed by commercial AI software requires extensive immersion in the field of application, initial development contracts call most clearly for treatment as service arrangements. If the finished program proves suitable for an entire class of users, however, subsequent sales might appear to involve a product. Ambiguity is inevitable where the uncertain intrinsic character of software diverts attention to its mode of supply for purposes of tort liability.

The question of causation raises a different set of issues—those chiefly related to the manner in which the computer program actually makes contact with the end user. The factors that evoke a clear liability rule for mass-marketed software likewise provide the strongest link between defect and injury. Simple sales transactions force the consumer to rely solely on the purchased item for effective performance, and thus, any harm suffered can be traced directly to improper program operation. It seems doubtful, however, that AI programs will reach consumers through such direct market channels any time soon. The simplest current reason is cost: AI software is enormously expensive to produce. A longer-range consideration is the likely reluctance of human experts to relinquish control over the provision of their expertise. Professions shielded by licensure requirements, for example, have shown themselves to be well equipped to defend against unauthorized practice. For the foreseeable future, then, the most likely role for many applied knowledge programs is as an aid to the human expert.

Although perhaps depriving AI developers from access to the consumer market, such restrictions also relieve developers of a great deal of potential liability. The law will not treat an appurtenant factor as a causal agent of injury unless it materially contributes to this injury; yet material contribution is precisely what is prohibited by restrictions on unauthorized practice. For example, if a physician were to attempt to lay blame on a diagnostic expert system for improper treatment, the physician would thereby admit to allowing the computer to perform as a doctor. The price of limiting the practice of a profession

For software developers, expenditures for debugging, design safety analysis, and quality control assurance are necessarily affected by the extent of perceived vulnerability to tort actions.

to a select group of peers is accepting complete responsibility for professional misjudgment.

Of course, if the source of the physician's error were indeed traceable to the expert system, the physician might sue the software developer to recover the money that must be payed to the injured patient. This possibility leads to the question: How should the law of product liability be applied to the creator of a device whose domain-specific capabilities might match or exceed those of human experts? The first step in any such lawsuit (whether based on strict liability or negligence) is to demonstrate that the product contains a defect. Should defectiveness be inferred from the mere fact of incorrect diagnosis? Human physicians are certainly not judged this harshly; their diagnoses must only be "reasonable." Separating programming errors from legitimate mistakes of judgment falling within professional discretion will indeed prove formidable.

A second, more practical obstacle facing this hypothetical physician is that a strict liability standard would probably not apply, regardless of whether the program is viewed as a product. Courts generally permit strict liability recovery only in actions for physical injury and property damage; economic loss is insufficient to trigger the doctrine. Unlike the hapless patient, the physician has personally suffered only financial impairment. The physician's tort suit, therefore, must be based on negligence. If the software developer has exercised reasonable care in debugging and packaging the product, including some statement warning of the system's limitations, the negligence burden might prove a difficult one to carry. The appropriate level of resources devoted to debugging efforts and the scope of the necessary warning depends on actual reliability and system design. For example, deep expert systems can generally be expected to deliver acceptable results over the entire useful range of the underlying causal model. Courts will undoubtedly expect greater vigilance from developers of shallow systems (or deep systems based on models that are not robust) simply as a consequence of the diminished reliability implied by program design.

To be sure, not all applications of AI techniques involve roles currently occupied by organized professions, nor must computer output actually touch consumers in order to affect their lives. The degree of contact necessary to trigger liability is once again determined by the nature of the commercial relationship. Personal interaction comprises an inherent feature of consultative service transactions, and a causal gap is probably inevitable unless the computer somehow communicates directly with the injured consumer; otherwise, the intervention of human judgment is likely to prove a sufficient superseding event to interrupt the nexus. Sales transactions, in contrast, are characteristically impersonal. The path of causation is likely to be far more direct if the computer's role involves assisting in the fabrication of a product. Any modicum of assistance not filtered through independent human oversight can furnish a link between computer operation and injury caused by defective manufacture. Hence, developers of computer-aided manufacturing (CAM) systems can expect increasing exposure to liability as their software assumes control over a greater portion of the manufacturing process.

Product design occupies a status somewhere between service and manufacture. Although the ultimate goal might be the production of a usable product, the design process involves early-stage development decisions of a far more basic and creative nature than those involved in automating production. The collaboration of computer-aided design programs with human engineers seems likely to persist for a much longer time than might be anticipated for CAM systems, which reduce design to actual practice, maintaining a greater opportunity for events that sever the chain of causation.

Increased trustworthiness inevitably results in heightened liability. As the role played by computers expands from passive assistant to independent practitioner, and as tasks delegated to computers begin to encompass a greater dimension of injury-producing activity, their owners will find themselves increasingly responsible for mishap traceable to improper operation.

Computers in the Courtroom

As AI research produces systems capable of real-time communication, computers can be expected to ascend from their present courtroom role as sidelined objects of legal controversy to active participation in the trial process itself. To lawyers, a computer serving as an expert witness offers unique potential advantages when compared to its human counterpart: quicker and more accurate responses to questions, greater capacity for immediately accessible domain-specific knowledge, and complete indifference to the charged atmosphere of a heated trial or the unnerving interrogation of a skilled attorney. For computer scientists, the specialized vocabulary and rules that govern judicial proceedings offer a relatively structured environment for developing knowledge-based systems capable of genuine interaction.

An initial issue confronting the introduction of computer testimony would be veracity: The law demands that witnesses give some assurance of truthfulness. The traditional oath-taking ceremony is no longer an indispensable feature of trial procedure, but all courts require that witnesses provide some form of affirmation backed by the threat of punishment for perjury. Could a machine furnish such an affirmation? No, not at present, because no contrivance yet devised is capable of anything resembling voluntary moral choice. Some human—programmer, user, or the litigant propounding the computer witness—must accept direct responsibility for the truth of testimonial output.

If the oath requirement presents some administrative inconvenience, restrictive rules of evidence stand as a far more formidable obstacle to the witness chair. As they currently stand, these rules bar today's computers from most courtroom affairs. Not only are current systems forbidden to act as witnesses, but their output often cannot even be introduced as substantive evidence. This impasse is created by a much-popularized but highly complex precept of evidentiary doctrine—the hearsay rule.

The rule owes its origin to the unfortunate judicial experience of Sir Walter Raleigh. In 1603, the legendary English adventurer, statesman, and author was accused of plotting to commit treason against King James I and forced to stand trial. The primary evidence introduced against Raleigh was a sworn statement made by his alleged co-conspirator, Lord Cobham, and a letter written by Cobham. Raleigh insisted that Cobham had recanted his statement and sought to question him directly. The prosecutor, Sir Edward Coke, refused to produce Cobham. Instead, he called a boat pilot who testified that while in Portugal he had been told, "Your king [James I] shall never be crowned for Don Cobham and Don Raleigh will cut his throat before he come to be crowned." Raleigh again protested, objecting that "this is the saying of some wild Jesuit or beggarly priest; but what proof is it against me?" Lord Coke replied, "It must perforce arise out of some preceding intelligence, and shows that your treason had wings." On this evidence, Raleigh was convicted and eventually beheaded.

Contemporaries considered the trial and its outcome cruelly unjust, and continued outrage ultimately resulted in the rule against the introduction of *hearsay,* that is, evidence whose source is not present in court for cross-examination. Unless a computer is capable of independently responding to questions, therefore, its output is unalterably hearsay and presumptively inadmissible.

The modern hearsay rule is not ironclad, however. Centuries of application have resulted in dissatisfaction with a doctrine that withholds so much highly informative evidence from the jury's consideration, and various exceptions have been introduced to mitigate its broad sweep. Computer output sometimes gains entry through one or more of these "loopholes." In particular, one widely recognized hearsay exception permits the introduction of routine business records; hence, the output of software that merely automates office functions often bypasses hearsay objection. Furthermore, the definition of hearsay is not all-encompassing: Only evidence introduced for the truth of its content falls within its terms. If a supposedly mute witness is heard to utter, "I can speak!" the content of these words is no more relevant than if this witness had said, "Banana." The ability to speak can be inferred from the act of speech itself. Similarly, a litigant is permitted to introduce computer output as proof that a computer was in fact employed (if such is an issue in the case), even if its content remains inadmissible.

What remains flatly barred is a computer's appearance as a witness, unless it possesses adequate communicative ability. In this regard, the demands of the hearsay rule are closely aligned with the objectives pursued by designers of natural language-processing systems: An acceptable witness must be able to understand questions posed by counsel for both sides and express responses in a manner graspable by judge and jury. The interactive processes implicated by this description can be decomposed into three constituent segments: (1) mechanical translation of speech signals into digital code, (2) interpretive analysis of the encoded text to derive meaningful information that can be acted upon, and (3) the formulation of coherent expressions of properly identified knowledge.[45]

Of these three segments, decoding speech signals is undoubtedly the least essential. Courts are often called on to accommodate the deaf and the mute; written questions and answers, shown to the witness and read aloud to the jury, satisfy courtroom protocol. Nonetheless, the ability to interact vocally with a witness is considered a highly valuable rhetorical asset by attorneys. Particularly when the subject matter is technical, juries are far less likely to fall asleep when observing a genuine interchange than when listening to a lecture delivered secondhand.

Researchers in this phase of speech-recognition technology have been more successful eliminating transcription errors at the intake stage than creating software to assign meaning to the correctly recognized words. Neither facet, however, has attained the level of verbal dexterity necessary to

The computer witness's inferential processes will need to be capable of operating at real-time speed in order to maintain the pace of courtroom dialogue.

facilitate truly interactive exchange. Template-matching techniques of word recognition must remain flexible enough to accommodate the acoustic patterns of a variety of speakers yet retain sufficient discriminatory power to decipher their messages accurately. Recognition difficulties are further heightened when speech consists of a continuous stream of utterances rather than isolated words. Disaggregating conjoined groups of words, whose individual pronunciation frequently depends on the surrounding phonetic signals, requires context-oriented analysis and the knowledge of allowed syntactic structures in order to cut through the ambiguity.[46]

"Understanding" the content of speech, which from an operational point of view might be defined as locating the appropriate internally represented knowledge in response to a question and fashioning an answer, represents the central focus of much current natural language research. Computer scientists are experimenting with a number of instructive models, but the greatest obstacles have been recognized for decades.[47] Vast amounts of subliteral inference contribute to the attribution of meaning in human communication, and an effective knowledge search requires precise characterization of this meaning. Contriving a satisfactory approach for bringing such soft concepts within the processing abilities of computer systems remains an elusive goal. Perhaps the recurring sequences of action and dialogue that take place in all legal proceedings can provide a basis for reducing the amount of world knowledge necessary for drawing the proper inferences. In particular, techniques based on scripts might prove useful as a means of formalizing courtroom discourse into patterns of expectable and, hence, more readily understandable verbal interchange.

Formulating a comprehensible response to a question once the appropriate knowledge is located requires more than simply translating it back into a natural language. In order to communicate effectively, additional features become necessary. A computer witness must be able to determine the level of detail and complexity appropriate for the audience (either autonomously or through external human calibration) and adjust its output accordingly. Answers perceived as unclear might need to be rephrased. Achieving the fluent command of language necessary to perform these tasks will demand a vocabulary of considerable richness and a highly sophisticated expressive capability. Once again, although numerous researchers have devised methods of generating limited semantic representations of linguistically unorganized data, systems possessing advanced cognitive faculties have yet to be developed.

These three phases of speech processing describe the ingredients necessary for basic voice communication; yet, while comprehensibility itself is a technologically formidable goal, it nevertheless describes only the surface of communicative ability. A satisfactory witness must be able to do more than interact. Especially where expertise in a particular field is claimed, the jury expects sound conclusions supported by adequate explanation. Attempts to implement the necessary reasoning processes will entail a great deal of structural planning at the level of knowledge representation. For purposes of witness testimony, the contours of required informational support are shaped largely by the proclivities of trial lawyers. Cross-examination is viewed by attorneys and judges as the acid test of witness credibility. When the witness is subjected to the cynical scrutiny of an experienced advocate, it is at least hoped that insincere assertions, erroneous convictions, and specious reasoning are exposed to the jury. Cross-examining attorneys rely on four principal avenues of investigative inquiry: witness qualification and experience, vulnerabilities in the witness's factual assumptions or scientific premises, the cogency of the ultimate opinion, and possible sources of bias.

The first of these areas is merely descriptive. The second requires some sort of recursive ability, such that conclusions can be traced to their origin; an attorney might wish to challenge the sufficiency of factual predicates or the respect accorded scholarly sources by the scientific community. The persuasiveness of an expert opinion is often tested by varying specific assumptions to determine whether alternative scenarios might produce indistinguishable conclusions. Thus, the computer witness's inferential processes will need to be capable of operating at real-time speed in order to maintain the pace of courtroom dialogue. The final aspect of the lawyer's inquisitory prerogative, a search for possible bias, engages some of the most difficult types of knowledge to implement in a computer's range of response. Bias is based on personal predilection. For an expert in a technical field, it can take such ill-defined forms as disdain for a particular scientific school of thought or a preferred method of approaching problems. The process of transferring expert knowledge to computer-based reasoning systems not only fuses individual experts' inclinations into the representational scheme, but can result in additional sources of bias as a consequence of programming design. The organization of a production system's control structure, for example, can involve conflict-resolution choices that tend to favor certain groups of rules.

These obscure sources of influence will remain beyond the reach of cross-examination unless the computer program is equipped to defend its own programming logic. If knowledge is simply gleaned from human sources and shuttled directly into a system of fixed rules, the computer will remain ignorant of the foundations underlying its expertise. True, this extraneous information might be painstakingly introduced into the computer's knowledge base solely to facilitate cross-examination, or perhaps, human experts could stand ready to provide supporting testimony. Either solution would render the computer a most ungainly witness. It appears that only those future "metaknowledgable" systems capable of developing and maintaining their own knowledge structures could feasibly maintain sufficient familiarity with chosen sources of intelligence. However, independent evaluation, selection, and absorption of raw information implies extraordinary computing power and currently represents a distant goal of AI research.

It is difficult to predict which phase of communicative capacity—comprehensibility or intellectual sufficiency—will prove the more intractable to develop. The measure of progress necessary for access to the courtroom is not solely a technological issue, however. As computers become recognized as authoritative sources of valuable knowledge, pressure will mount to permit greater use of their capabilities in judicial contests. Diminishing skepticism and the desire to produce the fairest verdicts will undoubtedly increase the tolerance for less than perfect communication skills, and prompt judges to allow sophisticated computer systems to take the stand.

Notes

The footnote style employed here is a specialized citation form peculiar to legal materials. This style has been developed to maintain consistency among jurisdictions, case reporters, and statutory citations. Its basic format is: <volume> <SOURCE> <first page of document>, <local citation>.

45. Voice synthesis would also comprise a useful, although not essential, element of the system. Because numerous commercial synthesizers capable of delivering perfectly acceptable speech output currently exist, it was not considered necessary to explore this aspect.

46. *See* 1 A. Barr and E. Feigenbaum, The Handbook of Artificial Intelligence 323 (1981).

47. An overview of current research directions can be found in Waldrop, *Natural Language Understanding,* Science, April 27, 1984, at 372. A recent example is described in Glasgow, *YANLI: A Powerful Natural Language Front-End Tool,* AI Magazine, Spring 1987, at 40.