Machine Help and Human Help in the Emerging Digital Library

Ronald J. Heckart

As the infrastructure of the digital library takes shape, a range of trends is gradually leading to the disappearance of human help in patron interactions. Taking the place of human help are ever more sophisticated types of machine help. This paper reviews these trends, posits a future scenario of almost total reliance on machine help, and explores implications and policy options for the profession.

A major change is under way in the profession. The change is no less than the disappearance of human involvement in patron assistance. No one in particular is directing the change; rather, it is driven by a confluence of technological and economic forces, and evolving notions of efficient service. In the not-too-distant future, it is likely that virtually all patron interactions will occur via networked computers. When help is needed, it will be provided online. A person may be available by appointment for one-on-one consultation, for group instruction in a classroom setting, or as a last resort after automated, online help has been exhausted, but even these options seem likely to be undermined by the inexorable trend toward ever more sophisticated—ever more virtual—forms of automated help and training.

In the logic of the digital library, the pertinent concept is user self-sufficiency. There is no overt intent to do away with human help as such; rather, the aim is to maximize unmediated access to information resources. Unmediated access is optimized when help is built in as an aspect of user-friendly design and as an explicit option users can invoke when needed. The aesthetic ideal is a smooth, elegant, and successful online transaction, free of interruptions caused by the need to consult printed manuals or human agents.

Large academic libraries, the current hotbeds of digital library development, are at the forefront of the general trend toward machine help, but other types of libraries cannot be far behind—indeed, the change can be viewed as an aspect of a broader transformation occurring in the commercial sector. The aim here is not to engage in a polemic about the evils or benefits of the change but, rather, to review the disparate trends that are converging to bring it about, to picture—albeit in a speculative way—what user self-sufficiency might be like in the digital library of the future, to examine the implications of the change for the profession, and to posit some policy alternatives.

Convergent Commercial Sector Trends

Library-oriented developments in online interactive help have their parallel in the

Ronald J. Heckart is a Librarian in the Institute of Governmental Studies Library at the University of California-Berkeley; e-mail: rheckart@library.berkeley.edu.
business world. Private industry is pouring millions of dollars into customer-support software research and development. An April 1996 Business Week article notes that the “market for customer-interaction software” is estimated to reach $3 billion in annual revenues by the year 2000. An array of high-tech companies is supplying ever more sophisticated interactive software to handle customer requests for information, customer complaints, and employee information needs and problems; and the platform for this software is increasingly the Internet and the corporate intranet.

Unmediated access is optimized when help is built in as an aspect of user-friendly design and as an explicit option users can invoke when needed.

Right now, corporations are implementing virtual help desks, in which an employee finds answers to questions by keying in a few key words on the corporate intranet. If negotiation is needed to refine the request, the employee is automatically prompted with questions. If an appropriate answer is not found, a customized help ticket is generated which might be routed to a particular department or person, or perhaps to the help functions of a more specialized software package or database. When the help system password is entered, information about the employee is supplied automatically and is used in handling the request. For example, the system might use the employee’s position in the company to grant access to some categories of information and not others, or to assign a priority level to the request. Because help software developers are standardizing the help ticket format, the help ticket could originate on the World Wide Web or an intranet, be invoked across a range of software packages, and still be recognized as a help ticket. Automatic transaction logs enable the company to analyze the performance of the help system.

Of special note are the increasingly powerful artificial intelligence features of help desk software, highlighted in this brief 1996 article:

AI-based help desk software deals with vague or unclear complaints. For example, a user may begin with the annoyingly broad challenge ‘it doesn’t print.’ Intelligent systems efficiently and consistently follow expert decision tree pathways to narrow the range of solution possibilities to the answer ‘connect the printer cable to the computer.’

AI natural language-based systems even allow users to enter plain English problems.

Neural networks, doing what they do best as pattern analyzers, allow the system to learn the users’ vocabulary. As the system solves problems, it automatically strengthens the basis for the resolution.

Help system databases automatically grow as new problems and solutions occur. Solutions in databases become part of the expert system IF/THEN knowledge base. The database grows dynamically. One user’s resolved riddle with Excel becomes available to others.

Surely, the future portends increasingly “intelligent” help desk software and expanding markets that encompass the kinds of information service that libraries provide. How much technology transfer actually will occur is difficult to say, but it seems certain that applications will not remain limited to hardware/software glitches and company information.

Convergent Academic Sector Trends
In academia, digitization and network connectivity are foundation stones for major changes in teaching, research, and library service. In the current technological ferment, the word digital is almost
synonymous with newness, power, and potentiality; and it is no surprise that
the phrase digital library has such cachet. The ascendancy of machine help is really
but one element of a panoply of changes in campus information technology. Giv­
ing special thrust to this ascendancy are some more particular trends.

One trend is the continuing innovation in educational technology, particularly in
the area of distance learning. Distance learners need access to information re­
sources and help with information-seeking just as resident learners do, and this
largely means reliance on machine-based systems. There is great interest in using
the Web as a platform for distance learning and in mechanisms to promote inter­
active course instruction. However, applications developed for distance learn­
ers are unlikely to remain limited to distance learners. Technological innovations
will surely find their way into the standard curriculum. Indeed, the distinctions
between distance learning and resident learning are likely to blur.

A second trend promoting the rise of machine help is the ongoing research ef­
fort in the library and information science (LIS) field to improve the user interface
in online catalogs, in part by incorporat­
ing expert systems into online catalog
searching. The goal “is to offer the informa­tion seeker assistance similar to that of
the human intermediary.” The migration
of online catalogs to the Web, with direct
links to full-text documents and images,
only heightens the need for more sophisti­
cated, “intelligent” online systems. There
is no expressed intention of doing away
with human help; rather, the goal is to im­
prove searching success and reduce reli­
ance on human intermediaries, although
in the long run the result may be the same.

A third notable trend indirectly facili­
tating the rise of machine help is the pat­
tern of academic library staffing over the
past decade or so. This pattern generally
has been one of steep decline. Pushed es­
pecially by fiscal constraints, academic
libraries have consolidated public service
points, reduced service hours, relied in­
creasingly on paraprofessionals and stu­
dents to provide service, and imple­
mented triage systems based on some

---

Distance learners need access to
information resources and help with
information-seeking just as resident
learners do, and this largely means
reliance on machine-based systems.

---

version of the “tiered” or “differentiated”
reference model. The next big step along
this trend line may be to provide little or
no in-person help at all. In this scenario,
machine help becomes a second-best al­
ternational or a restorative agent, depend­
ing on one’s perspective.

An Imagined Self-Help Scenario

From a user’s viewpoint, what sort of
future are these converging trends likely
to produce? Following is an imagined sce­
nario set in the 2010–2015 time frame. It is
informed by speculations in the library lit­
erature, and—quite frankly—by science
fiction. The protagonist in this scenario is
Alex, an undergraduate student.

When Alex enrolled at the university,
he was automatically assigned an avatar,
a graphical representation of himself, that
uniquely identifies him in the campus
computer system. Alex has customized
his avatar so that it actually looks like
him. Bundled with the avatar is a lot of
information about Alex. The avatar is his
electronic proxy for accessing campus
services, including the digital library.
Many campus services and many of
Alex’s classes have virtual analogs in the
campus computer system. Alex can visit
the registration office in person, but he
also can go there via his avatar. On his
computer, Alex clicks on the virtual cam­
pus gateway and moves his avatar to the
registration office. Once there, he finds a
virtual registration officer (actually another
avatar) and asks a question by typing it
on his keyboard. Alex’s question appears in a bubble of text above his avatar, and the answer is similarly displayed. In this instance, a true avatar, a graphical image with a real person behind it, responds to Alex’s question. In most such contacts, he deals with intelligent agents, complex programs of structured information that can be queried. The most sophisticated intelligent agents are represented as persons and sometimes are difficult to distinguish from avatars.

Alex is anxious about a research paper he must write for a seminar. Not only will the professor see it, but he also must post it to the class Web site for review by his fellow students. Alex moves his avatar to the digital library. He can stop at the central information desk and query the intelligent agents who reside there, or he can move directly to subject-based rooms where additional intelligent agents are ready to help him. The rooms are organized in a kind of hierarchy based on the Library of Congress classification scheme. Alex decides to go directly to the subject room that he feels most closely corresponds to his paper topic. There he queries an intelligent agent for information on the topic. The intelligent agent extracts keywords from Alex’s question, adds some synonyms, arranges them in a hierarchy, and sends the result back to Alex for review. Alex does some editing of the keyword list and adds some additional parameters: a date range and a format limitation to English-language books and articles. He then instructs the intelligent agent to perform a search. The initial search result is a sample of citations that Alex is asked to review. Alex can zoom in on the citations to get abstracts, tables of contents, and full text. He tags the citations he finds most relevant and deletes a few that seem too far afield. The intelligent agent takes this fine-tuning into account and performs a second, more complete search.

In the midst of Alex’s dealings with the intelligent agent, a classmate’s avatar strolls into the room. They exchange greetings and briefly discuss their paper topics. With this exception, the research process is a solitary activity which Alex conducts from his dormitory room with its excellent network connection. (Students who must dial in bemoan the slow response time.)

Alex reflects that his research is going well. But this was not always the case. He recalls some frustrating experiences which, in retrospect, he blames on his choice of topics. The topics had an unwieldy interdisciplinary cast, and his keywords were too general. Although Alex had made what he felt was a thorough perusal of online help features, he never felt he truly had a handle on the topics. He could have sought additional help by making appointments to visit a librarian avatar or a librarian in person, but he held back. He sensed an overriding presumption, true or not, that students should be able to succeed with intelligent agents and online help; failure most likely meant that he had done something wrong or exercised poor judgment in selecting topics.

Alex is now ready to begin building a hypercard of his search results. The hypercard can reside on his space in the campus computer system or on his own hard disk. As he reviews the material, he extracts parts of books and articles that he finds most relevant to his topic; rarely does he download a whole text. A citation is downloaded automatically with each extract, and Alex creates electronic notecards for each. At one point, when Alex clicks for a full-text display, he gets a warning message that he will be charged for viewing or downloading the document. Although the university heavily subsidizes the digital library, some documents are so expensive to provide that fees are assessed to discourage frivolous use and achieve partial cost recovery. Alex decides to download the citation with a notecard reminding him to review the citation again later. He will pay the fee if he is unable to locate enough
free (i.e., fully subsidized) material to meet his needs.

There is still a physical library with undigitized books in it, but Alex ignores it. The general attitude, shared by Alex, is, “If it’s not digitized by now, I can probably get by without it.” Besides, Alex would have to do his own scanning to include undigitized text in his hypercard, and this is too much trouble.

Alex begins organizing and expanding his electronic notecards, and gradually his paper takes shape. He resists the temptation merely to string together quoted passages, paraphrasing here and there, and he shuns outright plagiarism. The citations in his bibliography are hyperlinked to their sources, and the professor could easily check. (Alex knows that some students attempt the ruse of not citing works they plagiarize.) The composition module of the hypercard software prompts Alex through some fine points of editing and formatting.

Alex completes his paper and posts it to the class Web site without ever visiting the physical library or talking to a real library staff member.

Implications for the Profession
Taking into account the convergent trends making machine help ever more pervasive, and granting that some version of the Alex scenario may come to pass, it could be argued that doing away with human help in academic libraries would not in fact be a big step. Much of the basic infrastructure for automated self-service—sophisticated campus networks, widespread Internet connectivity, and computer-savvy clientele—is already in place, and help technology is improving rapidly. This view could be bolstered with the argument that automated self-service would be available anytime, anywhere, as long as the patron has access to a computer—a very considerable advantage.

There are other clear advantages to automated self-service. Human beings have good days and bad days; they get tired; any number of idiosyncrasies or distractions may affect their performance; they may not react well when the same question is asked again and again by a horde of students from the same class. But none of this matters to the machine. Moreover, machines have an anonymous impartiality that some patrons may prefer. Questions can be put to a machine which, when put to a person, may provoke feelings of embarrassment and diffidence. It can be argued further that libraries may as well take advantage of the cachet that computers and networks have with patrons and invest heavily in automated self-service—put resources where patrons are predisposed to use them.

An important, although controversial, goal of the profession has been to empower patrons to become self-reliant information seekers. Indeed, many librarians would argue that they are most successful when they impart information-seeking skills that patrons are able to put to productive and independent use. Another view, notably held by Herbert S. White, is that librarians succeed when they help patrons meet particular information needs and that librarians should be very circumspect in sharing their skills and expertise outside the profession, just as physicians and lawyers are. The concept of automated self-service complicates this debate in interesting ways.

On the one hand, automated self-service can be interpreted as increasing the patron’s independence. He or she is not dependent on direct human help but, rather, on the machine, which can be accessed anytime, anywhere, and which embodies professional skills and expertise in a user-friendly, natural language-based interface. True professional skills
and expertise are sequestered in the interface’s decision trees and neural networks, yet the patron makes productive use of them. On the other hand, dependence on the machine introduces its own vulnerabilities. For here the machine is an intricate, decentralized mass of networked protocols, data, and hardware. It can fail, sometimes catastrophically, leaving the patron utterly incapacitated. And there is risk in embedding professional skills and expertise “subconsciously” in the interface. Might they become so embedded over time that only handfuls of vendor-based programmers, cognitive engineers, and subject specialists would grasp the fundamentals of question negotiation and search strategy?12

Another issue that arises in the complex technological environment of automated self-service is equality of access.

It is pertinent here to note that the process of creating online help files is itself increasingly automated, as are cataloging and indexing functions, the ultimate goal being a seamless meshing of help, indexing, and data in “intelligent” information delivery systems. There is software, with suggestive names such as RoboHelp, that can be applied to other software to generate help files in Microsoft® Word or HTML format. Automatic indexers troll through the Internet enabling search engines such as Lycos and Alta Vista to retrieve relevant documents. The most sophisticated of these indexers analyze Web sites for keywords and phrases. Major efforts are under way to improve this indexing by creating standards for the description of Web documents. Using these standards, document authors and publishers will themselves input the descriptions (“metadata.”).13 It is as if book publishers embedded a MARC catalog record in the book itself, and the record self-loaded into the library catalog upon receipt of the book. Clifford Lynch foresees two levels of indexing in the electronic world.14 One level would be produced with input by human indexers and reviewers, and would be of the quality now associated with organized collections in libraries. This premium level would be fee based and presumably available in academic settings. The other level would be computer produced and look much like what Internet search engines offer now. This level, supported by advertising, would be free or almost free.

Even if a significant role remains for human indexers and reviewers, it is clear that a great deal of the activity that prepares electronic documents for access to users will happen automatically, with authors and publishers building standardized metadata elements into their documents to facilitate machine processing. This amounts to a remarkable diffusion of a kind of activity once thought to be the purview of librarians and libraries. Yet, one way to view this development is not so much as a shrinking of librarianship’s reach but, rather, as an extension of the work of organizing information resources for use. Machine indexing is providing some level of access to vast numbers of electronic documents which, in printed form, would be regarded as ephemeral and not suitable for retention in library collections, if collected at all. In any case, the larger point is that heretofore discrete, library-centered functions of making information resources available and providing assistance in their use are increasingly an integrated online process that is neither library centered nor oriented to the provision of help by librarians.

Another issue that arises in the complex technological environment of automated self-service is equality of access. Is it reasonable to assume that needed hardware, software, and network connections will be ubiquitous? Will not the
marketplace govern, such that the best, most up-to-date resources will be available to those who can pay the most? In academic settings, network access and other kinds of computer assistance, including access to software and fee-based databases, already are heavily subsidized. This is likely to continue, as is the decline in the cost of computer hardware—that cost currently is so low that personal computers already are de rigueur for most college students. So, however automated self-service evolves on a campus, it is likely that all students will have it. But will citizens at large have it? The answer depends on future public policy regarding information technology and to what extent social and economic disparities in American society are addressed in the years ahead.

But for many librarians these speculations dissolve in the face of one overriding conviction, expressed most succinctly by Cheryl LaGuardia: “People continue to need to work with other people in solving their information quandaries. Where do I get off saying this? Because I have been a reference librarian for 16 years.” In language less indignant, but no less incredulous, Jeannette A. Woodward says:

We see ourselves as teachers working alone or in tandem with the classroom instructor. Patrons, we say, especially undergraduates, rarely know what they want. They simply don’t understand enough about a topic to have a notion of where it fits in the world of ideas. We cannot imagine them giving coherent instructions to a know-bot when we must pull tentative, unformed thoughts from them, slowly and painfully. Then, with the patron’s full if reluctant participation, these fragile fragments must be woven into a productive search strategy.

The conviction is, in short, that no machine could ever achieve the intricate, interactive subtlety that many reference interviews require. Those who would rest comfortably with this conviction might be reminded of the many studies casting doubt on the performance of human-based reference service. If reference librarians find the right answer or point the patron to the right answer only about half the time, is the potential of the machine to be so firmly dismissed? A middle ground avoiding the excesses of both dismissive skepticism and unbridled optimism may be the prudent course here. It might be allowed that the machine will prove to be reasonably accurate with simple and mildly complex queries, but inaccurate with more complex queries, and queries made complex by the unfocused, confused way in which they are posed. Indeed, a sophisticated machine of the future might do a better job with simple and mildly complex queries than humans.

This middle ground also must accommodate values that go beyond accuracy in question answering. The patron may learn useful information-seeking skills in the course of a reference transaction or a training session, quite independently of getting the right answer to a particular question. And user satisfaction may be at least as important in assessing performance as accuracy in question answering. At present, these values seem best attuned to the interactive, interpersonal qualities of human help, and skeptics who doubt the machine’s potential might take comfort in this. But, again, the prudent course may be to keep an open mind and a watchful eye, particularly on developments in help desk software and online tutorials. For even if machine help is judged with this broader mix of values, it still may be found “good enough” from a policy standpoint. It may not need to be as good as, or better than, human help.
Policy Alternatives

In the real world of policy-making, decisions are based not only on rational assessment of alternatives, but also on considerations of the political environment, including prospects for financial support. Three broad policy alternatives are presented here. By no means do they encompass all the particular choices and rationales that could be brought to bear, but they are a useful launching point for a policy discussion.

If machine help proves to be one aspect of a future in which students work mostly at computers, mostly in isolation, a broad perspective seems essential.

The first alternative is to postpone substantive policy decisions and, instead, launch an explicit research agenda to pinpoint an appropriate dividing line between human and machine help. That line would not be static because of continuing improvements in machine help, and the research agenda would have to encompass the work of help software developers not currently offering library applications. Perhaps benchmarks could be developed so that as machine help met successively more stringent sets of criteria, human help could be curtailed correspondingly until such time as machine help reached the limit of its capacity. Perhaps that capacity is so great that human help would disappear, but perhaps not. This scenario would shift the whole matter into the realm of rational decision-making and give some objective basis for analyzing librarians’ concerns about machine help.

The second alternative is to make exclusive reliance on machine help a clear and deliberate policy goal. This choice could be based on the calculation that the gain in the accuracy rate with simple and mildly complex queries, coupled with the cost savings of doing away with human help altogether, outweighs the loss in the accuracy rate with more complex queries and concerns about broader values such as user satisfaction. Most librarians would object to this trade-off, but administrators and budget officers might not. Another way to arrive at this same option is to make the political calculation that, whatever one’s personal reservations, the realities of funding and resource allocation are such that the future lies with the digital library and its corollary, user self-sufficiency, and that to continue to assert the need for human help undermines the profession by making it seem stodgy and backward. Still another way to arrive at this option is the calculation that if machine help has won the day, librarians’ best course is to put users’ interests first and make machine help as effective as it can be. This would mean greatly expanded efforts to work directly with software developers and vendors to improve machine help.

The third alternative is a more effective and affirmatively stated version of the status quo. The status quo is mostly remarks in passing about the convergent trends discussed above, with defiant assertions of the continuing need for human help. A more positive approach would allow that some forms of machine help may be efficacious but would affirm that assistance with complex or poorly framed queries is best provided by human beings. It would note the risks of relying on complex technologies without human backup and would incorporate sociological and psychological perspectives as well. If machine help proves to be one aspect of a future in which students work mostly at computers, mostly in isolation, a broad perspective seems essential. One potential problem such a perspective might bring into focus is user disinclination to seek human help if it is retained in some form as a last resort option. The availability of such help might reassure the librarian, but the inference that the user might draw is that, normally, one ought to succeed with
machine help and that turning to human help is evidence of ineptitude.

Conclusion

Unless issues of machine help and human help are given more explicit attention in the profession, it is likely that human help will gradually disappear from the scene without anyone quite realizing that it has happened. There is great irony here, for the ultimate goal of the programmers and cognitive engineers is to replicate human help at its best—to replicate something that we already have (at least sometimes) and are now abandoning. It is not so much that a fundamental reorientation is needed; rather, it is that library trends need to be reexamined with a help-centered sensibility, supplemented with an attentiveness to developments in the help desk software industry. As these shifts of focus occur, the sorts of policy options discussed here will emerge in greater clarity, and librarians will be better poised to be leaders rather than followers in their rapidly evolving future.

Notes

8. The author follows the practice of many futurologists in setting the time frame far enough ahead that he will not be proved wrong quickly.
10. Avatar technology is developing rapidly as an enhancement of Internet chat lines. At the CNET Web site (http://www.cnet.com/Content/Reviews/Compare/Chat/), Bronwyn Fryer and Dave Trobridge evaluate the latest avatar-based chat services, “where you take on the identity of a graphical character in a near-hallucinatory 3D world.” They give the highest marks to Worlds Chat (http://www.worlds.net/wcg/), which combines “wonderful 3D graphics with a full gallery of fun avatars to choose from, easy-to-use messaging features, and great online help.”
13. There are two leading sets of standards. One, originated for U.S. federal government documents and services, is called Government Information Locator Service, or GILS (see http://www.usgs.gov/gils/gilsCopy.html). The other is called the Dublin Core (see http://purl.org/metadata/dublin_core).


17. “Unobtrusive” studies, in which proxies pose as patrons and ask questions of reference staff, have consistently found accuracy rates in the 50 percent range. For a concise summary of these studies, see Patricia Hults, “Reference Evaluation: An Overview,” Reference Librarian 38 (1992): 151–73.