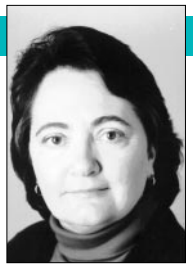

Is AI going mainstream at last? A look inside Microsoft Research

By Sara Reese Hedberg
sara@hedberg.com

For years, many in the AI and intelligent systems field have yearned to grab the brass ring of the mass market. Longing for broad deployment of AI into mainstream computing, some have even aspired to become the "Microsoft of the AI world."

However, after AI's 15 minutes of fame in the mid-80s, analysts grew tired of waiting for AI to hit the big time and tucked it into a "promising technology" niche. Although AI has continued to make significant strides in theory, technology, and even deployed applications, its impact has yet to reach its full potential and profitability.

There are many feathers in AI's cap—the most recent is the IBM Deep Blue supercomputer's defeat of world chess champion, Garry Kasparov. In large corporations, thousands of systems that include AI save hundreds of millions of dollars annually in finance, factories, and offices. Household, electronic, and office appliances, for example, are a little smarter because bits of AI help them autotune their performance and self-diagnose malfunctions. A suite of software tools that include rule-based AI chunks run retail chain stores such as Mrs. Fields Cookies. Robots now serve meals to hospital patients, clean up toxic dumps, and explore Mars. The examples of AI-enabled intelligent systems are many and varied, including a range of technologies such as heuristics, neural networks, genetic algorithms, natural language processing, and case-based reasoning. For the most part, they are limited to solving a particular kind of problem—such as scheduling or process control—in a particular market segment—manufacturing, for example, or insurance underwriting.

Microsoft is beginning to change all that. Steve Ballmer, Microsoft's executive vice president of sales and support, and Bill Gates' buddy from Harvard days, addressed the National Conference on Artificial Intelligence in Seattle in 1994 (AAAI-94), declaring that AI is very important to Microsoft. He gave demonstrations of Tip Wizard in Excel and Liz in the Magic School Bus, little intelligent agents that help users understand new programs. Ballmer admitted that these were simple agents—only the start of the intelligence Microsoft hopes to build into future products.

Bill labs

Bill Gates has a vision of the computer of the future—an intelligent assistant that talks, listens, sees, and learns, yet is simple and intuitive to use.

To achieve this vision, Gates began assembling a computer brain trust called Microsoft Research. First he hired Nathan Myhrvold, who graduated from UCLA at 19, and from Princeton at 23 with a PhD in mathematical and theoretical physics. Myhrvold left his fellowship under Stephen Hawking at Cambridge University to start up a software business that Microsoft bought in 1986, and he joined Microsoft.

In 1991, he began building the Research group now, led by some of the field's legends:

- Gordon Bell, father of Digital's VAX minicomputer;
- Chuck Thacker, coinventor of Ethernet and chief designer of the Xerox PARC Alto computer, which spawned the Mac;
- Charles Simonyi, who designed the Alto machine's software, which inspired the revolutionary Mac interface;
- Gary Starkweather, inventor of the laser printer; and
- Alvy Ray Smith, who cofounded Pixar, the company that made Toy Story.

Microsoft's brain trust of researchers, now totaling 250, continues to grow. The company plans to hire another 400 researchers in the next three years.

While most corporations are slashing research budgets, Microsoft is pumping its up. The company spent between $20 and $30 million in 1996, and its 1998 budget is estimated between $100 and $200 million for the Research group alone.

"The role of research within Microsoft," explains Jim Kajiya, assistant director of the group, "is to create strategic technologies for Microsoft that form the core for new products. Microsoft is committed to developing cutting-edge products. To do this, we need ready access to cutting-edge ideas and research." Microsoft Research has the pull to build such a brain trust, offering attractive salaries, lucrative stock options, and all the market muscle of the software giant.

Decision Theory and Adaptive Systems

Bill labs' distinguishing feature is that it gives researchers the opportunity to see their ideas, theories, and algorithms quickly become products used by millions. In 1993, this possibility pulled three AI scientists from their Palo Alto startup, Knowledge Industries. As the principals of Knowledge Industries, Jack Breese, David Heckerman, and Eric Horvitz were invited to Microsoft for what they thought was a sales call. They were recruited.

The three licensed their technology, based on Bayesian probability theory, to Microsoft and joined Research, starting the Decision Theory and Adaptive Systems group (*http://research.microsoft.com/research/dtg/main.htm*).

With the shipping of the Answer Wizard in Office 95, DTAS work began to appear in Microsoft products, and the possibility that attracted the trio began to become a

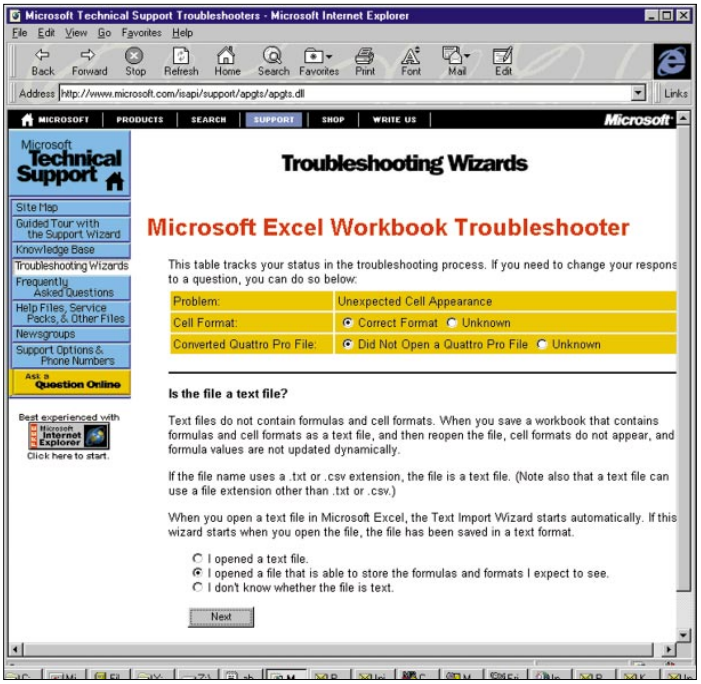

Figure 1. Microsoft Troubleshooting Wizards.

reality. Since then, threads of DTAS work have been woven into Office 97, Web-based support services, and MSNBC. The group has 20 people and continues to grow. Their work spans a number of areas, including responsive user interfaces, troubleshooting, data mining, intelligent computer-resource allocation, text retrieval, and causal discovery. While the group maintains strong Bayesian underpinnings, other theories and techniques are also under development for building flexible and adaptive software systems sensitive to user preferences and actions in large-scale distributed environments.

Microsoft Research encourages these scientists to collaborate and cross-pollinate with other groups throughout the company. The DTAS group works with a number of other core technology groups, such as Natural Language, Database, Speech, Vision, Advanced Development Tools, and OS, and product groups such as Office, SiteServer, NT-C, MSN, MSNBC, WebTV, and Visual Studio.

Troubleshooters with a twist

Diagnostic troubleshooting of system failures was one of the first application areas where AI techniques, notably rule-based heuristic systems, proved their worth in the late 70s and early 80s. From manufacturing to medical care, a host of such applications have been deployed—venture it to say they have become commonplace. For those who follow trends in the AI field, hearing about yet another troubleshooting system is something of a ho-hum experience.

But some of the DTAS work puts a noteworthy twist on the world of troubleshooting applications. What began as discussions between Jack Breese (DTAS) and Product Support Services in late 1993 led to a proof-of-concept system called Aladdin, a Bayesian-based help desk assistant. With the popularization of the Internet, the target shifted to self-help applications in an attempt to reduce calls to the Microsoft staff.

The result is some 70 Web-based user self-help troubleshooters that have already been deployed by Microsoft Technical Support Services, and they are adding more (see Figure 1). Using DTAS's Microsoft Belief Network tool (see "A Bayesian briefing" sidebar), TSS built the Bayesian network and did the system engineering,

A Bayesian briefing

The world of statistics has two main categories: classical and Bayesian. The main difference between the two varieties," according to Decision Theory and Adaptive Systems Senior Researcher David Heckerman, "is that Bayesian statistics embraces prior knowledge whereas classical statistics does not. That is, Bayesian statistics provides an explicit mechanism for combining prior knowledge (for example, the knowledge of an expert) with data." It gets its name from the Reverend Thomas Bayes, who in 1763 wrote an essay setting forth a mathematical formula to calculate probabilities among several causally linked variables.

Bayesian theory can handle the contingencies, probabilities, and uncertainties inherent in many situations. It affords a means to automatically generate predictions and decisions with incomplete and uncertain data by reasoning about causes, problems, processes, and tasks. Bayesian probability theory has been applied to troubleshooting equipment failures, medical diagnostic expert systems, data mining, Web data filtering, space shuttle control, and responsive user interfaces, to name a few.

The DTAS group has developed the Microsoft Belief Network (MSBN), a Bayesian tool used internally at Microsoft and available to noncommercial research and academic organizations at no charge (*http://research.microsoft.com/msbn/default.htm*). There are now more than 600 external-to-Microsoft users worldwide.

MSBN is intended for building, testing, and teaching about belief networks. The tool enables the creation, assessment, and evaluation of Bayesian belief networks (see Figure A). Bayesian networks are graphical

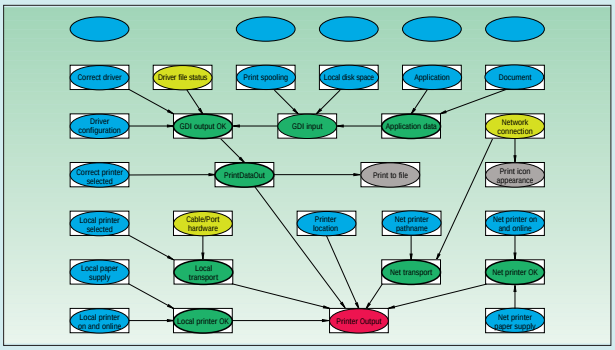

Figure A. Sample Bayesian belief network.

models of relationships and dependencies between variables, including probabilities. Users can represent "belief networks" of complex cause-and-effect relationships between nodes, weighting them to express how nodes affect one another. One of the beauties of Bayesian networks is that, rather than being deterministic, they can handle the notion of "might." With influence diagrams, specific actions can be tied to costs and benefits based on the belief network. Thus, the system can infer the most reasonable course of action, given the network of relationships, probabilities, and trade-offs inherent in the situation. Written in 32-bit Visual Basic, MSBN comes with a Windows dynamic-link library and several sample belief networks and runs on Windows 95 and Windows NT (3.51 or newer).

such as the user interface, memory management, and html links (*http://support.microsoft.com/support/tshoot/default.asp*).

Although there has been no rigorous statistical analysis of this approach's effectiveness, the support organization seems quite satisfied. The proof is in plans to build more Wizards and to shrink-wrap existing ones into Windows 98 and NT 5.

The electronic butler

Imagine the day when your desktop PC is working behind the scenes to keep an accurate model of your interests, tastes, predilections, needs, and actions. It uses your "model" to provide custom-tailored assistance in everything from writing a letter to performing Web searches—in short, for everything you do on your computer. A number of such projects are underway among AI researchers world-wide.

Eric Horvitz, a DTAS senior researcher, has developed Lumière—an intelligent agent named after the butler in Disney's film *Beauty and the Beast* (a French candlestick with a Maurice Chevalier accent). "The art of butlering," explains Horvitz, "is a subtle skill, centering on the difficult task of aiding someone by careful observation—and assisting without interfering."

Lumière is a Bayesian-based system for automating more responsive interfaces based on user modeling (*http://research.microsoft.com/~horvitz/lum.htm*). In one study, Excel neophytes worked with the spreadsheet software in one room, while in another, experts monitored the novices' actions on networked PCs. The experts typed messages to the users suggesting ways they could assist. The information from such studies was encoded into Lumière, which includes a Bayesian network model of user activity and needs (see Figure 2) and continuous online monitoring of user actions to infer butlering tasks (see Figure 3). For example, if a user opens a Word document and types "Dear Mrs. Green," the system infers the user wants to write a letter and offers assistance.

A shallow version of this technology was shipped as the Office Assistant with Office 97. Users can choose from various icons as the front end for Office Assistant, such as a paperclip, a picture of Einstein, or my own favorite, William Shakespeare (see Figure 4). The Office group chose to only include a simple Lumière capability for the release because of other considerations,

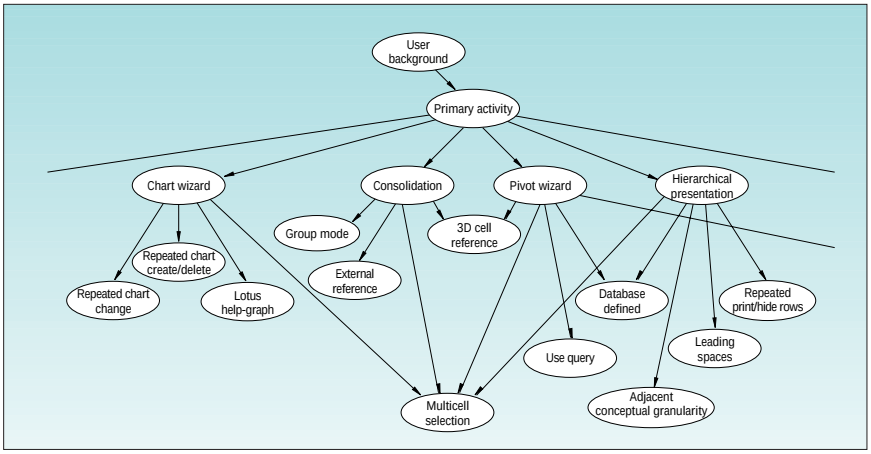

Figure 2. A portion of the Lumière Bayesian network user model.

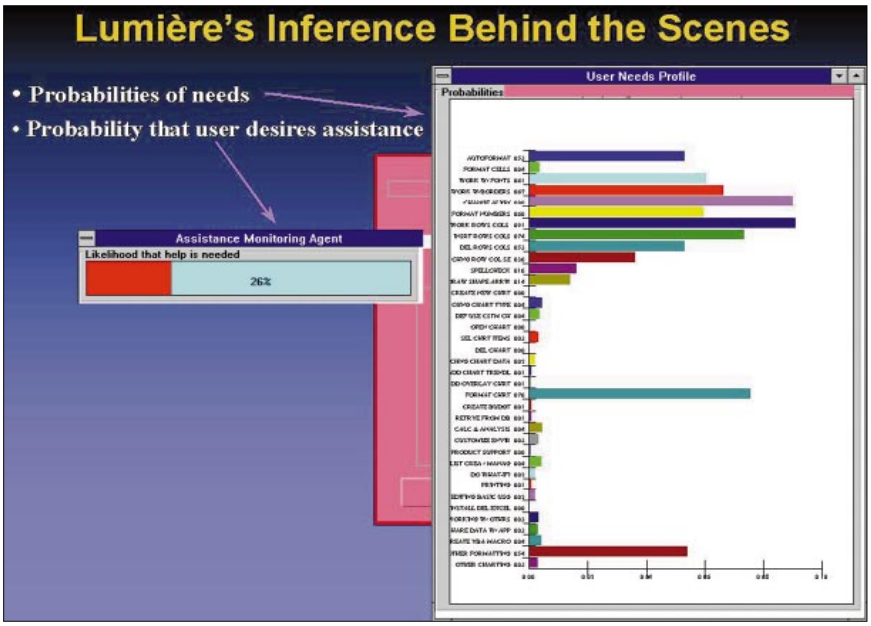

Figure 3. Lumière's Inference Behind the Scenes.

such as needing to focus significant resources on Internet-related features.

For researchers lured by the possibility of getting their ideas into the mainstream market, Office Assistant certainly hit the mark. Office 97 is the fastest-selling business application in PC history, according to Microsoft. In just over a year, it has sold more than 30 million licenses—at an average rate of 70,000 per day.

"The intelligent assistant in Office is not the end of the line for development of advanced help for end users," notes Kajiya. "I believe we will continue to develop more. Microsoft systems are very complex, with lots of features. A major problem for the Office group is to make people aware of all the features." He sees Office Assistant as just the start of intelligent help systems that present the right things at the right time. "Intelligent help will accelerate," he predicts. "It will appear fairly quickly in a primitive form, and it will evolve."

Indeed, Bayesian networks can become quite complex in their modeling abilities. An early Bayesian-based medical diagnosis system that Horvitz crafted at Microsoft illustrates this point (see Figure 5).

Getting personal on the Web

Similar technology has been used for other DTAS projects. In one Web-based project, for instance, the system watches users in the background to learn behavior patterns. The goal is to provide personalized services. "If we have data about lots of users, and your pattern is similar to other users, then the system can suggest that here are other sites on microsoft.com where others like you go," explains Jack Breese. "We have algorithms

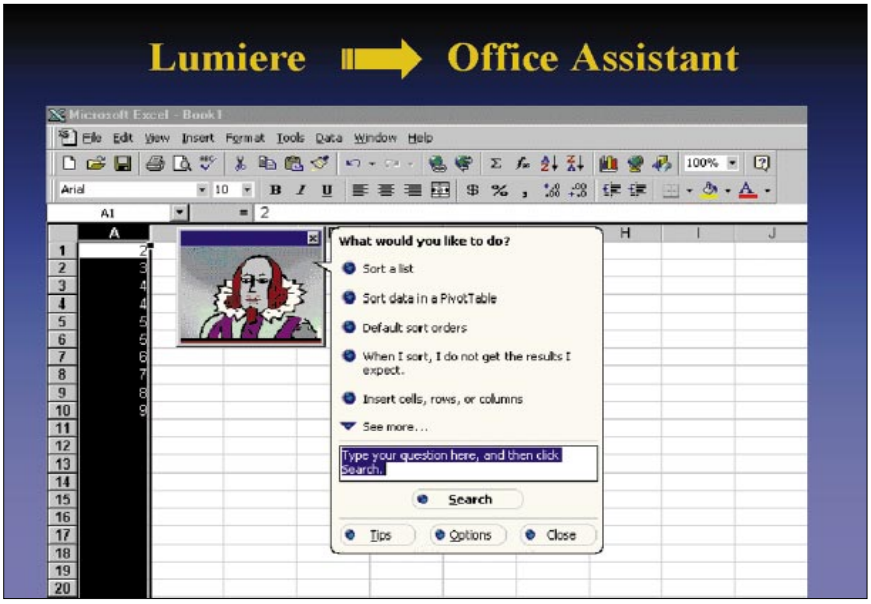

Figure 4. Lumière instantiated as the Office Assistant.

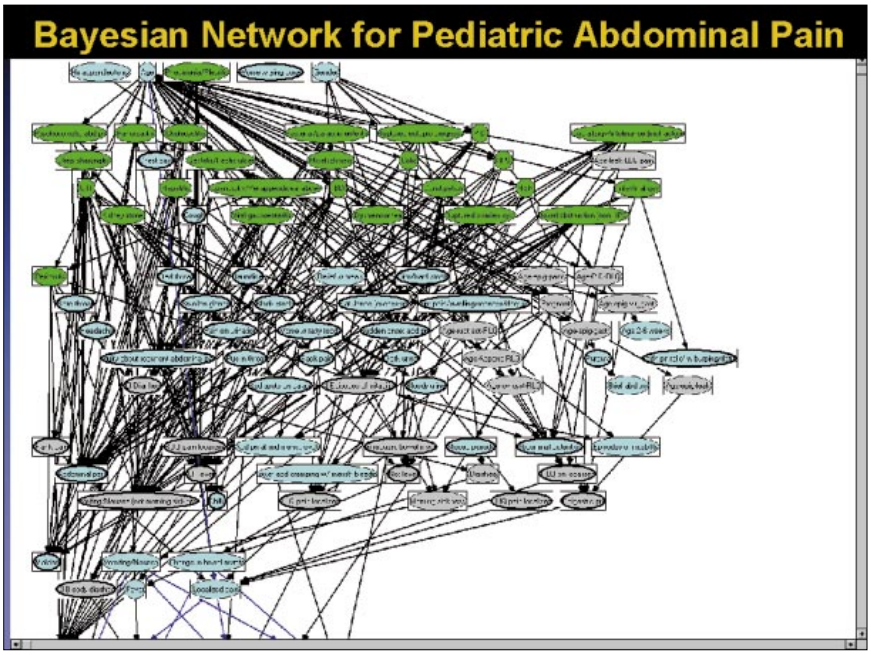

Figure 5. Bayesian medical diagnostic system.

to do this." Some algorithms are Bayesian-based; others are not.

Imagine a user logging on to a Web site for the first time. She answers some questions about her preferences, which the system can use, coupled with information it has already gathered about similar users (called *clustering*), to suggest other portions of the site. The ability to personalize online content is already present on the MSNBC Web site, visited by more than 300,000 people each day. At the bottom of the front page at *http://msnbc.com/news/* is an option to "Personalize this Page," which launches the process.

It is not a real stretch to imagine that this technique would enable cross-selling of items in electronic commerce. Indeed, IntelligentCrossSell is about to be rolled out in Site Server 3.0, Commerce edition (see Figure 6). Senior researcher Usama Fayyad worked with others in DTAS to build this feature.

Plus data mining

To its suite of megatalent, DTAS has recently added Fayyad, one of the leading researchers in the emerging field of knowledge discovery and data mining. Before joining Microsoft, Fayyad spearheaded the Caltech Jet Propulsion Lab Skicat data-mining application.

These days, Fayyad is working on other issues, including data-mining operators for clustering and classification of data. Although some database companies now offer SQL extensions called data-mining operators, Fayyad believes these fall far short of what is really needed to efficiently mine large data stores. Currently, the standard approach to get a cluster count of database records segmented by age groups, for example, is to load all the data into memory and then do a successively refined sort that may take a number of passes over the same record. This is a highly expensive computing process. Using machine learning, Bayesian networks, and other techniques, Fayyad and crew are aiming at more computationally efficient approaches that are scalable for large data sets, such as a "get the counts" classification operator that would traverse the database only once, tabulating the count into a table.

DTAS's small but growing data-mining group is doing seminal work. Research agreements are already in place with major companies in fields such as pharmaceuticals to test some of the new operators and approaches on live data. As this work matures, it will likely feed into one of Microsoft's new targets—the enterprise market. Microsoft now has only about a two-percent market share, but with data mining folded into its product offerings, that could all change. For now, Fayyad is rather circumspect about when and where such operators will see the light of day as products.

Intelligent computer resource allocation

Besides work in the computationally efficient mining of data stores, the DTAS group is also investigating other ways to improve computational resource monitoring and allocation. Some work in Web browsing focuses on how to use idle CPU time to fetch related information in the background. Members also are working with the Millennium project, part of the next-generation systems effort at Microsoft. The goals for the Millennium distributed operating system include self-tuning and self-configuration, which interleave closely with the adaptive-systems focus of DTAS.

Further reading

S. Hamilton, "Inside Microsoft Research," *Computer*, Vol. 31, No. 1, Jan. 1998, pp. 51–58.

L. Helm, "Improbable Inspiration," *Los Angeles Times*, Oct. 28, 1996; http://www.hugin.dk/lat-bn.html.

R.E. Stross and A.H. Morre, "Mr. Gates Builds His Brain Trust," *Fortune*, Dec. 8, 1997; http://mouth.pathfinder. com/fortune/1997/971208/mic.html.

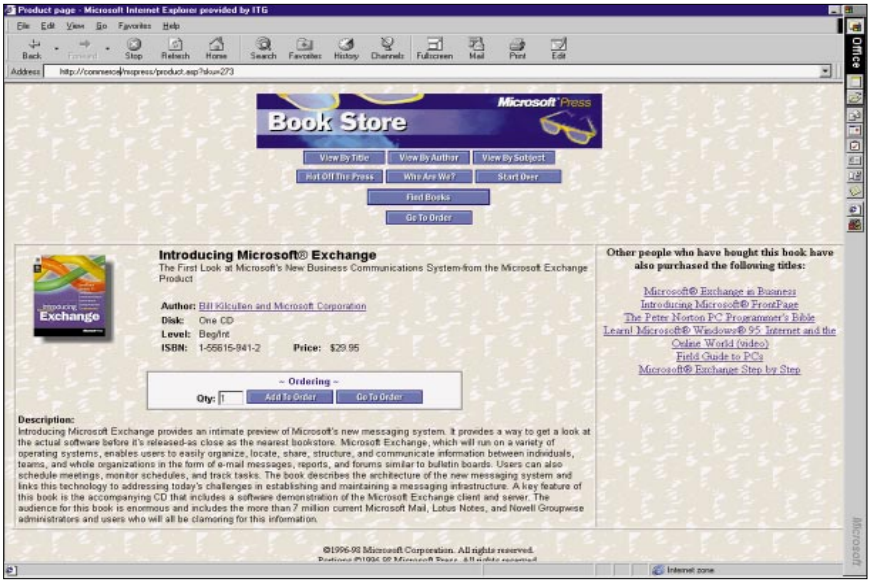

Figure 6. A Microsoft online bookstore visitor has shown an interest in a book about MS Exchange. On the right side of the screen is a list of related books generated by the system.

Along these lines, DTAS is investigating intelligent, dynamic graphics rendering. Horvitz's Qualia (a word used in neurophysiology to signify the way things appear visually to people) is a research project studying how to balance image quality in an environment bounded by limited computational resources. The group has run human studies to understand what parts of a screen's image are most important to perception. Based on the studies, they are finding that not all parts of a graphic image need be displayed with the same degree of resolution. Modulated granularity could reduce the computational costs, dynamically finding the balance between the costs and benefits of computing resources and human perception.

Eric Horvitz describes it this way: "We're looking at the resources for continual computation. We want to make systems that are always anxious. We want them always anxious about how best to serve the user—to take CPU idle time and magnify it using probability theory." This research area spans many of the core groups at Microsoft, such as operating systems, multimedia, and networking. Although it is too early to talk about how this work will impact future products, it is a sure bet that such adaptive-systems work will soon see the light of day.

Advancing theory

The DTAS group's work is an impressive blend of advancing the state of intelligent systems theory and building algorithms and tools that are going into products. Some work has found its way into Windows, Web services, Office 97, and MSNBC; others will appear soon, such as IntelligentCross-Sell in Site Server 3.0, Commerce edition. Other work likely has a longer time frame, such as some of David Heckerman's causal discovery work. Essentially, he is refining a theory for inferring and learning about causes and effects. Such work might prove useful in areas such as building causal models of what users are doing to provide more effective assistance.

New to the group in the summer of 1997 is Susan Dumais, formerly at Bell Labs. She is concentrating on intelligent information-retrieval research interleaved with some of the user-model work. The goal is to create a system as good as a human librarian at providing information that meets each user's specific needs. Currently, when two different people ask the same question, they get the same answer. On the desktop, an adaptive system can use the user's context garnered from recent e-mail messages, questions, and other activities to guide the Web search and retrieval process.

Dumais and Heckerman, using probabilistic-tools (Bayesian networks) methods for information retrieval and machine learning, are making headway in their research. They have tested some of their work, for instance, by running it on a standard test set of Reuters data. It was able to automatically classify the large data set with 91–92% accuracy. They are also applying machine-learning techniques to help determine what to do with an item once it is classified—something more useful than simply piling it on your desktop.

Internal groups at Microsoft handling lots of text, such as product-support and Intenet sites, as well as product groups, might start to use this work.

Moving fast

Moving from initial innovation to product can happen fast at Microsoft. Such has been the experience of the DTAS group. "The Bayesian system will be in the next version of NT for any application to use," according to Kajiya. Eventually we could see generic collaborative filtering techniques in Web-authoring tools applied to a host of different tasks. The Intelligent Assistant is just the tip of the iceberg in terms of what we might see in intelligent user assistance. Other groups in Microsoft Research are working on natural language, speech recognition, development tools, and more.

Microsoft has the brain trust and market position to bring to the market software for smarter desktops. Has Microsoft grabbed the AI brass ring? Are the headwaters of intelligent systems for mainstream computing really on the Microsoft campus in Redmond? If so, for those of us who get exasperated by how much time we waste because of our stupid PCs, this could be good news indeed.

Acknowledgments

I thank Jack Breese, Susan Dumais, Usama Fayyad, David Heckerman, Eric Horvitz, and Jim Kajiya of Microsoft, and Shelby Barnes of Waggener Edstrom for all their assistance in researching this article.

Sara Reese Hedberg is a freelance writer and regular columnist for both *IEEE Intelligent Systems* and *IEEE Concurrency*. She has written extensively about trends in AI and intelligent systems for more than a dozen years.