In: IEEE Intelligent Systems, Vol. 21, No.5, pp. 14-18, Sep./Oct. 2006

Intelligent Interactive Entertainment Grand Challenges

Mark Maybury, *MITRE*

Oliviero Stock, *Trentino Cultural Institute Center for Scientific and Technological Research*

Wolfgang Wahlster, *German Research Center for Artificial Intelligence*

Advances in AI and human-computer interaction offer seemingly endless opportunity to enhance traditional forms of entertainment and support the creation of new ones. However, several fundamental scientific and technical impediments must be overcome to achieve these promises.

Webster's defines entertainment as "something diverting or engaging" or "a performance." IT occurs in a broad range of domains including games and toys, the fine arts, movies and radio, sports, travel, and education. Figure 1 shows some examples from the digital world, including virtual games, interactive arts such as audience-guided movies, augmented sports (performance or viewing), virtual and nonvirtual travel guides, and computer-based tutors. In each case, there are various ways to support users—for example, by guiding the material and information, by enhancing the users' performance, by expertly critiquing their performance, or by providing online help or guidance. However, providing support requires knowledge not only about the particular entertainment domain (and its associated entities, relationships, tasks, and activities) but also about interaction competence.

Intelligence in games

AI encompasses multiple facilities:

- Perception includes seeing, hearing, and feeling.
- Cognition addresses analysis, planning, and forecasting.
- Communication embraces speech, language, gesture, and discourse.
- Manipulation includes creation, movement, and destruction of entities.
- User and agent modeling enables representation and reasoning about beliefs, preferences, goals, and plans.
- Learning addresses abstraction, generalization, failure analysis, and adaptation to novel situations.

Each area offers valuable capabilities to digital entertainment. Figure 2 considers how these facilities apply to intelligent games. For example, perceptual processing such as highlighting objects or events of interest can enhance user experience and memory. The enhanced realism of intelligent behavior or strategy representations can enrich cognition. Human-language technologies promise more effective communication between users and the game as well as more natural and believable game-character communication. Kinesthetic models can more effectively represent the virtual world and improve the ability to reason about and manipulate embodied characters. Agent languages support the design and reasoning of more natural artificial characters. Finally, user models enable the personalization of information and presentations to a changing set of user goals, beliefs, and preferences.[1]

Intelligent interaction

Human-computer interaction is an AI application area that addresses aspects of an intelligent interface: As figure 3 shows, these aspects include multimodal, tailored, interactive, cooperative, mixed-initiative, and model-based interactivity.

Multimodal interfaces support multimedia (for example, spoken or written language, gesture, gaze, and graphics) and multimodal input analysis and output generation (for example, audio, visual, and haptic) as well as multimodal discourse processing (for example, within and across modal reference and discourse). Cooperative interfaces help users detect and correct misattributions or misconceptions, avoid undesirable events and states, and complete dele-

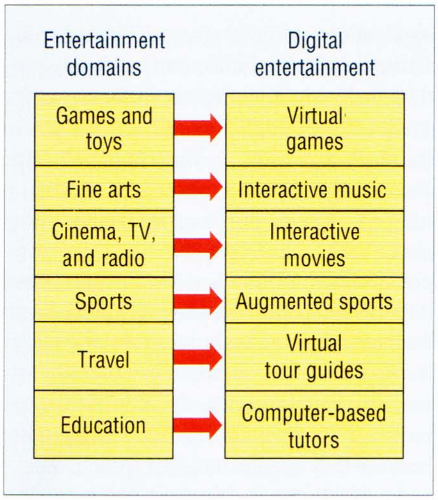

Figure 1. Digital entertainment domains and examples.

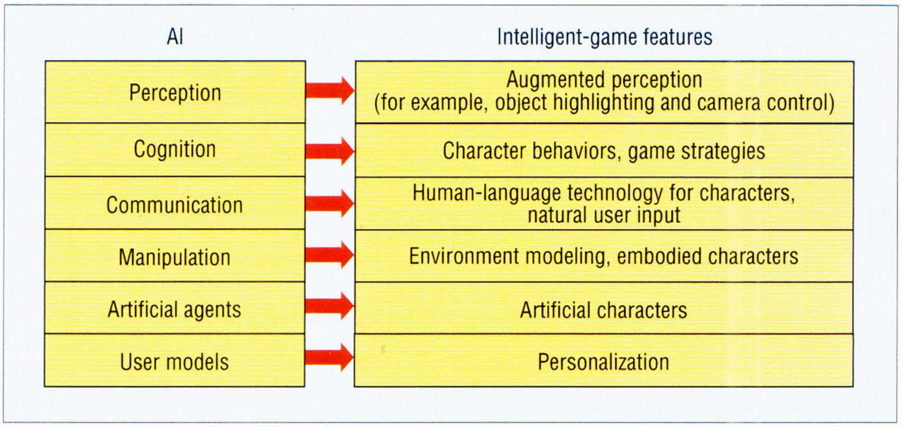

Figure 2. AI facilities featured in intelligent games.

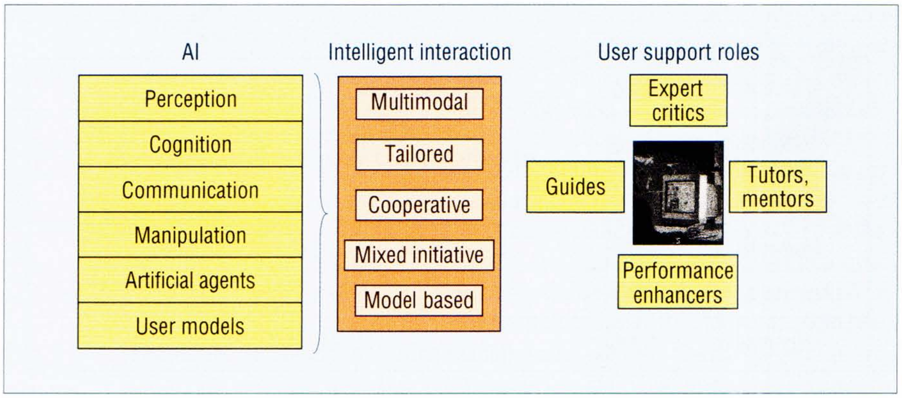

Figure 3. Intelligent interaction supports entertainment users in many roles

gated tasks. Mixed-initiative interfaces let the user or system at any point assume the leading role in the interaction; they require the flexibility to lead a dialogue and respond naturally when a user barges in. Tailored interaction utilizes task, user, or discourse models to customize the interface; the models are based, for example, on task needs or user preferences. Model-based interfaces incorporate declarative models of the user, domain, task, style guides, and interface elements to support the semiautomated design and redesign of the interface itself.

Intelligent interfaces can support entertainment users in various roles, as illustrated at the right in figure 3. They can guide users through entertainment materials and activities, serve as mentors or tutors to help users perform activities, act as expert critics of user plans or designs, and actually enhance individual or team performance. This interaction could range from a simple direct-manipulation interface (for example, a recommended menu of actions) to an embodied conversational agent. Embodied characters afford many instantiations, such as a virtual mentor character in an interactive game, an augmented-reality tour guide in an art museum, coplayers in a virtual orchestra, or a personal tutor in a virtual classroom.

Domain challenges

Given the role of intelligence in entertainment generally and in the interface in particular, we next consider possible challenges in each entertainment domain.

Interactive games

Just as humans compete in Olympic games that include both individual events and decathlons that assess overall athletic achievement, digital interactive games encompass both individual challenges and ones that draw upon multiple intelligences.

Interactive, intelligent games have a long history, including specialized expert game players such as IBM's Deep Blue, which beat world chess champion Gerry Kasparov in 1997. After that controversial success, based on special hardware, the field continued to advance by extending the use of AI. Current leading chess programs are probably stronger than Deep Blue, even if they run on normal PCs. Researchers are also developing many other game experts, including a system that can beat humans in solving real crossword puzzles.[2] This system requires specific natural-language and information-retrieval capabilities.

Besides specialized programs, AI is addressing general game playing. In GGP, programs must learn rules about games as opposed to just perform within a single game's predefined rule set. The 2005 AAAI GGP competition aimed to "test the abilities of general game playing systems by comparing their performance on a variety of games."[3] Players are told nothing about the games to be played before the competition. At the beginning of each game, they receive its rules electronically in the GGP Game Description Language (http://games.stanford.edu/language.html). Systems compete against one another for a US$10,000 prize.

Measuring a game's learning competence is one challenge. Assessing its enjoyability or realism from the user's perspective poses other challenges. These might include creating games that let the user and system collaborate to predict some quasi real-world outcome, such as the price of a barrel of oil in a Sims-like world or the most effective immigration policy. These games might measure success by how fast the system-user team achieved the goal or by their proximity to the optimal solution within some specified time period.

Interactive digital arts

Figure 4 illustrates a range of ways in which intelligence and interaction can enhance the digital arts.

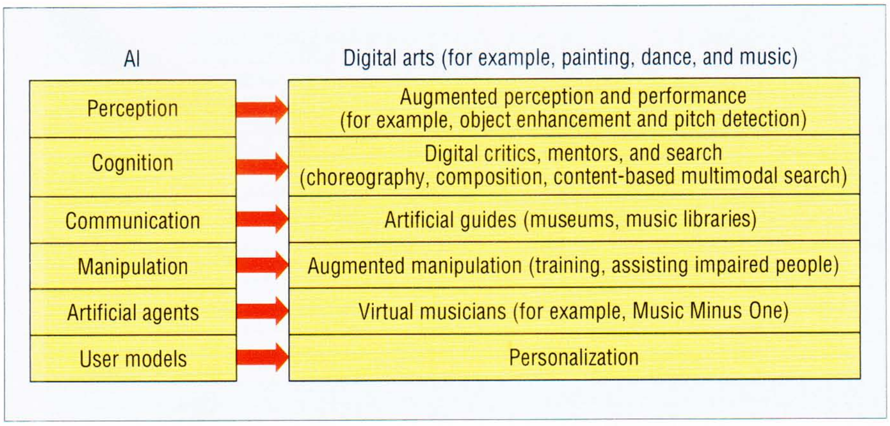

Figure 4. Interactive-entertainment challenges in the digital arts.

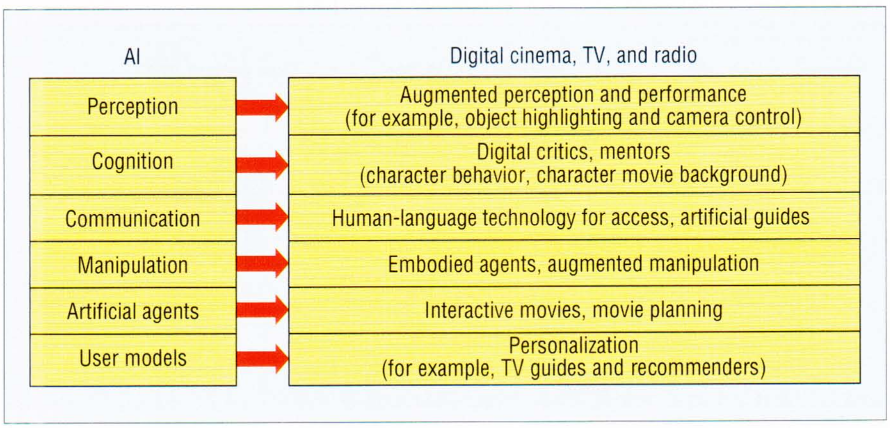

Figure 5. Challenges in the digital cinema, TV, and radio domain.

Visual, acoustic, or haptic devices can augment both perception and performance. Such devices could facilitate performance for people who are perceptually challenged or just make it easier for average users. A primitive example is visual highlighting of the keys on interactive-training keyboards. An acoustic example is automatic pitch detection and correction in Karaoke machines to assist tone-deaf or off-key singers by readjusting the music pitch to match the singer.

Graphic design, music composition, and dance choreography all require detailed knowledge of presentation elements, their properties, the events that prompt them, and how to create those events. Performance critics and mentors can use this information to support learning and performance. Researchers have already demonstrated content-based multimedia search to discover artifacts such as pictures, sounds, and movements.[4]

Beyond simple search, embodied conversational guides could provide access to large artifact collections such as those in museums or museum archives. We could use models of individual interests and preferences to personalize these guides in terms of content, sequence, and presentation.

Embodied agents could serve as partners for dance or musical performances. Popular static forms of this idea already exist, such as the Music Minus One (www.musicminusone.com) series. A more interactive, reactive form could include virtual musicians or dancers that would react to a user's actions and possibly improvise their responses.

One challenge would be to have a novice artisan perform with assistance (either alone or with other artisans). A panel of human judges could then assess the quality and style of the resultant performance, which might feature drawing, painting, dancing, or playing an instrument. In another form, the interactivity could support coordination among various artists—for example, choreography for orchestral or band song composition. A challenge task might be to support novel designs or competition against an expert human (a kind of artistic Turing test).

Interactive cinema, TV, and radio

Figure 5 illustrates a range of intelligence applications to digital cinema, TV, and radio. Directors often enhance movies during production by highlighting particular objects or events (for example, the red-dressed girl in the black-and-white movie *Schindler's List*). An intelligent design critic can comment on movie scripts or storyboards. By designing choices in plot lines, producers can give viewers an active role in a story's progression, although outcomes are generally predefined. Human-language technology or manipulation-enhancement methods can provide natural multimodal access to cinema, TV, and radio for all. These artifacts are challenging because they contain mixes of speech, music, and natural sounds. Directors and producers commonly use nonlinear digital editors today. Increasingly, tools incorporate content-based search, which can include model-based understanding of broadcast news,[5] movies, and videotaped meetings. User models already support recommender systems for movies, books, and music.

Example interactive-movie challenge tasks include finding key music or video segments or answering questions about characters' behaviors or events in a movie. We could measure performance in terms of both task accuracy and timeliness or perhaps even user enjoyment.

Interactive digital sports

Like many forms of entertainment, users can enjoy sports both as performers and observers. Sports observation is typically in person or via television. The examples from the previous section are relevant in the case of TV. In the case of performance, the implementations can be either physical or virtual. Virtual interactive games are very popular in many sports such as football, baseball, soccer, and hockey. Users can play as individuals or team managers, participating directly in play and strategy; they can also modify player parameters, select situations, and even manage camera control. AI plays a key role in digital sports to include player, team, and crowd behaviors; human-language technology for character interaction; as well as augmented perception by manipulating camera angles and replay speeds to achieve entertaining effects.

In addition to virtual sports, a variety of augmented-reality sports have become popular. Examples include paintball and laser tag. These sports combine sensors and effectors with computers to, for example, count the number of hits or dynamically establish team memberships, thus enhancing the situations and reactions to situations that players experience individually and collectively.

Interactive sport challenges arise naturally in the games themselves, of course. Because games can measure both individual and team performance, they provide an interesting set of possible success measures. These can include not only raw scores or hit/miss ratios and more fine-grained individual-skill assessments (such as posture and movement in time and space) but also the quality of strategy or the degree of distribution or success in coordinated team activity.

Interactive travel

Virtual or augmented travel has become a popular preparation for or alternative to physical travel. Virtual tour guides similar to those used in museums can help plan any itinerary. They might interactively tailor itineraries according to the interests of a particular traveler or travel group. They might also support human-language interaction by answering questions about historic sites. These might eventually be mobile robots, but they could just as well be disembodied agents that interact with users via handheld mobile platforms. In Boston, visitors to Zoo New England (the Franklin Park or Stone Zoos), Minute Man National Historical Park, or the city can receive information about current attractions interactively via their cell phones in their preferred language and voice, using technology developed by Spatial Adventures (www.spatialadventures.com).

Intelligent agents might do much more in support of travel, such as planning itineraries optimal to user's desiderata, scheduling routes sensitive to local weather conditions, or simply negotiating transportation purchases. GPS and enabled wireless devices with onboard digital maps (including 3D and subsurface applications) enable diverse possibilities.

An interactive travel challenge could measure performance of an intelligent agent or agent-supported user to perform a travel-planning task within a given time to maximize some set of measures such as a combination of economy, excitement, and novelty. Or you could simply have a scavenger hunt driven by interesting locations.

Interactive education

So-called "edutainment" already incorporates interesting interactive elements. Reading, writing, and drawing assistants help novices learn basic skills. Some incorporate, for example, speech processing (including accent models) so that young readers learning their native or a foreign language receive immediate and individualized performance feedback. Incorporating interactive games seems to enhance user motivation. In the future, incorporating question-answering systems could help provide more tailored mentoring.[6,7]

Animated pedagogical agents in interactive learning environments enable effective situated learning.[8] For example, researchers have shown that a qualitative simulation of the cardiac cycle in a virtual patient provides a safe, effective 3D virtual emergency-room training environment for student physicians.[9]

Challenges are evident in tasks such as training and testing users both with and without intelligent interactive training. Metrics for evaluating this work include error rates and time to learn. A problem's complexity level or team-based problem-solving could also establish interesting evaluation opportunities.

> **When artificial agents become part of our environment, we expect communication to have the same basic properties as human communication.**

Home and society challenges

While perhaps not obvious as an entertainment form, home and other social environments can be natural spaces for interactive entertainment. For example, some people cook for fun; an interactive assistant could do everything from help with online shopping and menu planning to suggesting music or games to play at a dinner party. Or it could have specialized knowledge for home repair or gardening. Interactive pets can provide companionship. A finding agent could keep track of a child's whereabouts or more mundane things such as the remote control or a lost sock. A social-analysis agent might help discover social opportunities based on social-network analysis or simply remind the user of important dates such as birthdays and anniversaries.[10]

Interactive-home challenges range from keeping a house clean to advising on key tasks. Measuring success in this environment includes the number of errors, cost savings, and increased user satisfaction.

Communication as entertainment

Besides its applications in artistic forms, communication in general can be an entertaining activity. When artificial agents become part of our environment, we expect their communication to have the same basic properties as human communication. For example, they might exhibit humor, storytelling capabilities, or even produce pictures or video clips to share important information.

The entertaining properties of communication can play a practical role in many situations. Consider humor in advertising, where it first helps get the audience's attention and then helps them to remember the message. Or think of the essential role of entertainment in children's education.

If a storytelling system can engage the listener, for example, it can provide a good companion to a senior person and convince her of the importance of taking medicine and doing some exercise. Such a system can also provide live sport commentary.

Entertaining communication is beginning to deliver some initial prototypes. Humor is a complex capability to reproduce, and some experts have considered it to be "AI complete." It's nevertheless possible to model some types of humor production and to aim at implementing this capability in computational systems. Some prototypes can produce expressions limited in humor typology but meant to work in unrestricted domains.[11,12] Besides humor production, work has begun on humor recognition, including research reported in this issue.[13]

There's also some work on automatic emotional storytelling, where the artificial character evokes empathy in the audience.[14]

Early results are available for automatic video production, especially in the simplified context that doesn't involve shooting in the real world but starts instead from 2D existing images.[15] Robot-based systems that produce good-quality pictures of events constitute another promising avenue.

All these themes open substantial challenges for the future. They require more insight into emotion, cognition, and computation. For example, what's the nature of a surprise in communication and how can we maintain attention?

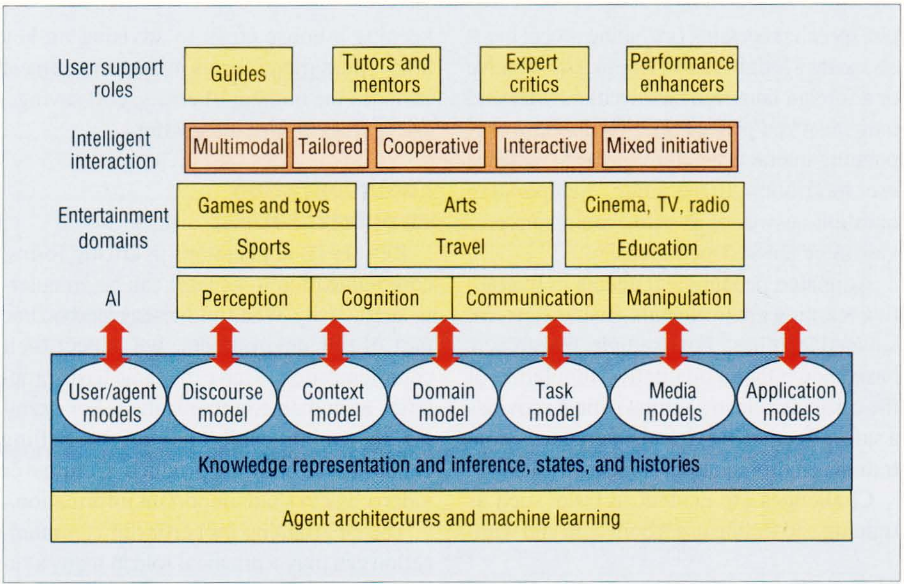

Figure 6. A digital-entertainment architecture.

With so many entertainment domains and applications, one challenge to applying multiple intelligences to enhance them will be creating reusable platforms that support multiple entertainment activities and interactive requirements. Figure 6 presents a digital-entertainment architecture that supports models of the user, discourse, context, domain, media, and application as well as the underlying open agent architectures and machine learning. Models of emotions such as fear, joy, and surprise will become important both for virtual characters and for understanding human-user input, both satisfactions and frustrations. Figure 6 only begins to suggest the complexity and opportunity richness of interactive intelligent entertainment. ■

References

1. A. Merlino and M. Maybury, "An Empirical Study of the Optimal Presentation of Multimedia Summaries of Broadcast News," *Automated Text Summarization*, I. Mani and M. Maybury, eds., MIT Press, 1999, pp. 391–401.
2. M. Ernandes, G. Angelini, and M. Gori, "WebCrow: A WEB-Based System for CROssWord Solving," *Proc. 20th Nat'l Conf. Artificial Intelligence* (AAAI 05), AAAI Press, 2005, pp. 1412–1417.
3. M. Genesereth, N. Love, and B. Pell, "General Game Playing: Overview of the AAAI Competition," *AI Magazine*, vol. 26, no. 2, 2005, pp. 62–72.
4. M.T. Maybury, ed., *Intelligent Multimedia Information Retrieval*, AAAI/MIT Press, 1997.
5. M.T. Maybury, "Broadcast News Understanding and Navigation," *Proc. Innovative Applications of Artificial Intelligence*, AAAI Press, 2003, pp. 117–122.
6. M. Light and M. Maybury, "Personalized Multimedia Information Access: Ask Questions, Get Personalized Answers," *Comm. ACM*, vol. 45, no. 5, 2002, pp. 54–59.
7. M.T. Maybury, ed., *New Directions in Question Answering*, AAAI/MIT Press, 2004.
8. W.L. Johnson, J.W. Rickel, and J.C. Lester, "Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments," *Int'l J. Artificial Intelligence in Education*, vol. 11, 2000, pp. 47–78.
9. M. Cavazza and A. Simo, "A Virtual Patient Based on Qualitative Simulation," *Proc. 8th Int'l Conf. Intelligent User Interfaces* (IUI 03), ACM Press, 2003, pp. 19–25.
10. M. Maybury, R. D'Amore, and D. House, "Expert Finding for Collaborative Virtual Environments," *Comm. ACM*, vol. 14, no. 12, 2001, pp. 55–56.
11. K. Binsted and G. Ritchie, "Computational Rules for Punning Riddles," *Humor*, vol. 10, no. 1, 1997, pp. 25–75.
12. O. Stock and C. Strapparava, "Getting Serious with the Development of Computational Humor," *Proc 18th Int'l Joint Conf. Artificial Intelligence* (IJCAI 03), AAAI Press, 2003, pp. 59–64.
13. R. Mihalcea and C. Strapparava, "Technologies That Make You Smile: Adding Humor to Text-Based Applications," *IEEE Intelligent Systems*, vol. 21, no. 5, 2006, pp. 33–39.
14. M.Y. Lim, R. Aylett, and C.M. Jones, "Affective Guide with Attitude," *Proc. 1st Int'l Conf. Affective Computing and Intelligent Interaction* (ACII 05), LNCS 3784, Springer, 2005, pp. 772–779.
15. C. Callaway et al., "Automatic Cinematography and Multilingual NLG for Generating Video Documentaries," *Artificial Intelligence*, vol. 165. no. 1, 2005, pp. 57–89.

The Authors

Mark Maybury is executive director of MITRE's Information Technology Division. His research interests include intelligent information systems, analytic tools, collaborative environments, distributed computing, and knowledge management. He received his PhD in artificial intelligence from Cambridge University. Contact him at Information Technology Div., MITRE, 202 Burlington Rd., Bedford, MA 01730; maybury@mitre.org.

Oliviero Stock is a senior fellow at the Trentino Cultural Institute's Center for Scientific and Technological Research (ITC-IRST). His research interests include AI, natural language processing, intelligent user interfaces, and cognitive technologies. He received his education at the University of Florence and the University of Pisa. Contact him at ITC-IRST, Via Sommarive 18, 38050 Povo, Trento, Italy; stock@itc.it.

Wolfgang Wahlster is the director and CEO of the German Research Center for Artificial Intelligence (DKFI) and a professor of computer science at Saarland University in Saarbrücken, Germany. His research interests center on language technology and intelligent user interfaces. He received his doctoral degree in computer science at the University of Hamburg. Contact him at DFKI GmbH, Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany; wolfgang.wahlster@dfki.de.