Informing and Promoting Technical Career Choices: Roadmaps for the Future

Dan M. Davis, Julianne M. Nordhagen & John J. Lazzeroni
University of Southern California
Los Angeles, California
dmdavis@acm.org & {nordhage, lazzeron}@usc.edu

Frederica J. Stassi
Science Education Analyst
Lompoc, California
frjstassi@gmail.com

ABSTRACT

The issue addressed by this paper is that of providing the nation sufficient technically qualified personnel to maintain national economic strength and defense ascendency. Considered first is the current need and prospective shortfall in Science, Technology, Engineering and Math (STEM) personnel. The genesis of the false impressions that many students have about STEM professions is discussed. These are often based on media and popular culture; the negative impacts are outlined. Two approaches are reported and analyzed: 1) the use of on-line "conversational avatars" to act as mentors for individuals choosing careers and 2) staging public gatherings for young people to visit with practicing STEM personnel. In the case of the on-line solution, there is a brief discussion of the University of Southern California (USC)/Office of Naval Research MentorPal project. Its focus is on target populations that have little access to STEM professionals to act as mentors due to geographical remoteness or Socio Economic Status (SES) disadvantage. Then, the bulk of the paper describes the authors' participation in a STEM fair organized by Professor Andrea Armani of USC. It is provided so students in the Los Angeles Basin can have direct contact with vital, engaging and accessible researchers. The authors cover how important decisions were made about: what the ultimate goals were, who should present science at the fair, which groups to target for invitations, ways to convey the science careers via posters, how to quantify attendance, and in what ways to best characterize the impact on the attendees. Preliminary data is adduced concerning all of these aspects of the effort. The materials are presented in a way that should enable other institutions to more effectively consider implementing programs of their own. Future research plans are described and evaluated.

ABOUT THE AUTHORS

Dan M. Davis is active as a consultant at the Institute for Creative Technologies, University of Southern California (USC), focusing on large-scale DoD simulations and avatar uses. Prior to retirement, he was the Director of the JESPP project at USC for a decade. As the Assistant Director of Advanced Computing Research at Caltech, he ran Synthetic Forces Express, helping bring High Performance Computing to DoD simulations. He also served as a Director at the Maui High Performance Computing Center and at the Jet Propulsion Laboratory and Martin Marietta. He was the Chairman of the Coalition of Academic Supercomputing Centers and has taught at the undergraduate and graduate levels. As early as 1971, Dan was writing programs in FORTRAN on one of Seymour Cray's CDC 6500's. He saw duty in Vietnam as a USMC Cryptologist and retired as a Commander, U.S.N.R. He received B.A. and J.D. degrees from the University of Colorado in Boulder.

Julianne M. Nordhagen is an Ensign in the U.S. Navy and is attached to the NROTC unit at USC, where she is assigned to collaborate on research at the Institute for Creative Technologies (ICT). As an Industrial and Systems Engineering (ISE) student, Julianne cultivated the analytical and technical skills required for this project, including optimizing the performance of large scale systems and applying computational methods for scheduling and forecasting. She has professional experience using integrated planning software like SAP and NetSuite ERP. She was selected to be one of the mentors for ICT's MentorPal project for the Office of Naval Research. Next April, she will begin her Naval Aviator training at the Naval Air Technical Training Center in Pensacola Florida. Ensign Nordhagen earned an ISE B.S. degree from the University of Southern California.

John J. Lazzeroni is an Electrical Engineering student in the Viterbi School of Engineering at the University of Southern California. He is currently serving as a Research Assistant and Project Manger in the Armani Research Laboratory there. His research interest flow in many directions, but he is interested in the need to attract more young people to the technical fields. His duties in the lab include organizing the EngX STEM Fair to try to provide a background for students of all Socio Economic Statuses. He is a Midshipman, First Class in NROTC at USC, where he has been selected as the Battalion Commander for the midshipmen. MIDN 1/C Lazzeroni was commissioned in the US Navy as an Ensign this spring. ENS Lazzeroni received orders to Nuclear Power training, after he received a B.S. in Electrical Engineering from USC in June.

Frederica J. Stassi, Ed.D. is a Science Education Analyst, working in the Central Coast of California. Her background includes research for the National Science Foundation in which she was funded to study pedagogies and efficacy in U.S. Science museums. This research involved museums from the East Coast to Oʻahu in Hawaiʻi. Her doctoral research was conducted under the guidance of Professor William McComas and focused on the development of science standards for the State of California. She received a BA degree from Tabor college, as well as an M.A. Degree in music performance and an Ed.D., both from the University of Southern California.

Informing and Promoting Technical Career Choices: Roadmaps for the Future

Dan M. Davis, Julianne M. Nordhagen & John J. Lazzeroni
University of Southern California
Los Angeles, California
dmdavis@acm.org & {nordhage, lazzeron}@usc.edu

Frederica J. Stassi
Science Education Analyst
Lompoc, California
frjstassi@gmail.com

INTRODUCTION

Survival of any entity is based on its ability to maintain its strength. There is an often cited and much lamented shortfall in technical personnel in the United States (BLS, 2014). The members of this community are sometimes referred to as Science, Technology, Engineering, and Math (STEM) professionals. The purported cause for this has been examined by many authors (Smith & Anderson, 2014). These causes include the negative images of technical people in the media, the lack of awareness of the range of STEM careers, and the mistaken notions of the actual working environments of both technical and non-technical professional fields (Maoldomhnaigh & Hunt, 1988). In many cases this is exacerbated by the lack of easy student accessibility to technical personnel, especially among students who are younger, who come from lower socio-economic-status (SES) households, and those living in communities that are geographically remote from centers of technical activity *e.g.*, rural areas (Lederman, 1992).

This paper reports on an activity where two complementary types of STEM career selection outreach were combined. Specifically, a prototype computer-generated and on-line capable STEM career mentor technology called MentorPal was displayed in a different setting, *i.e*. as a booth during a to a person-to-person outreach event on a major university campus in an urban setting. In this case the University of Southern California (USC) is located in downtown Los Angeles. Through both the use of the avatar technology and by a live professional explaining their experience in technology, researchers sought to increase the students' awareness of the range of activities that STEM personnel perform. They sought to augment the students' views by illuminating them about the wide variety of personalities of technical personnel and about the paths that lead to various technical careers. It was a putatively unique opportunity to compare and contrast the computer approach with the live mentoring approach.

This paper first lays out the virtual conversation technology, reviewing some of the background research that undergirds MentorPal and setting out how mentors are recorded and converted into computer-based interactive mentors. Next, it discusses insights from the STEM fair event, in terms of student reactions and interaction dynamics in that context. Finally, the paper discusses research in progress and potential future directions that might increase the effectiveness of this approach.

GENISIS OF THIS PROJECT

To those who have not had to face the challenges of its hurdles, creating a virtual human may seem as simple as remodeling a human using Computer Generated Imagery (CGI.) It turns out that it takes significant study and effort to implement a Virtual Conversation and the process can consume considerable computing power. The range of skills and the application of technologies that go into the creation of a virtual conversation with lifelike abilities include Natural Language Processing (NLP), machine learning, Virtual Reality (VR), CGI (if animation is involved.), and social stimulation of humans by computer-generated interactions. NLP will be the main focus of this discussion, though the same argument concerning the limits of virtual humans can be made with several of the other components. Natural Language Processing is comprised of the decomposition of language to allow the computer to do useful communications (Chowdhury, 2003). Recent developments in NLP have made significant advances, including breaking down sentences into: parts-of-speech tags, chunks, entity tags, semantic roles, similar words, and the grammatical and semantic elements of a sentence that generate meaning (Collobert & Weston, 2008).

The MentorPal project draws from earlier approaches using some of the same underlying natural language dialog technology; ICT's contribution is known as NPCEditor (Leuski & Traum, 2011), which specializes in question-answering agents. NPCEditor is one of the two methods used by MentorPal in selecting the best video clip to respond to the users questions. One of the first implementations of the NPCEditor technology was an exhibit at the Boston Museum of Science called The Twins, where visitors interacted with dialog-based virtual agents who could answer questions about computer science and about how they worked (Swartout *et al*., 2010). Later research examined the use of the underlying dialog technology to reproduce the experiences of Holocaust survivors, by producing recorded video clips of the living Holocaust victims to be used as responses, in an attempt to make the conversations more personal than using animated avatars (Traum, *et al*., 2015). These programs relied on a range of interactions and social dynamics. Often they included a guide who would work direct a group of visitors in taking turns asking questions to which the program would answer. In the program under consideration here, MentorPal has been designed to focus on lower cost, program designer control, more manageable question sets and more mentor-like dialog. This "faster-cheaper" approach was intended to enhance adoption of this method by others.

This work was all based on the previous and extensive literature on mentoring relying on Artificial Intelligence (AI). The focus of MentorPal is on emulating the experience of an informational interview, such as the kind a student would have with a mentor in a counseling office or at a career fair. This can be contrasted with counseling via systems that are designed to help participants working on a research project. An example of that would be AutoMentor (Wang *et al.*, 2013). It can also be compared to intelligent mentor agents which gave support to metacognition as part of an open learner model (Dimitrova & Berna, 2016). On the other hand, MentorPal is not intended to build skills. Even among mentoring agents which also handle question asking and familiarization, *e.g.* the SimCoach system (Rizzo *et al.*, 2011), MentorPal is distinct due to its focus on subjective experiences. Instead of a general description of STEM careers or fields, it is intended to help learners identify a mentor whose experiences resonate with them and that enables them to explore a more realistic vision of a career.

One of the issues of consequence is that many users, especially the target teen-age demographic for this research, have now become familiar with on-line conversational agents. Most of these other agents are simple question/answer operations, a kind of an automated FAQ. However, earlier work with NPCEditor and similar systems do not speak to some of the critical MentorPal project issues, especially that many of the users really do not know what to ask when evaluating a career, have already made simplistic uninformed choices, and must be "primed" before using the system or prompted later when they cannot formulate germane questions on their own. As this is an emerging technology, careful monitoring of user reaction is indicated. Therefore, this STEM fair exhibit offers a different perspective on interactions with this kind of system. In particular, for the STEM fair it was not certain if students would find the system novel, due to their greater familiarity with engaging virtual assistants. Prior testing with MentorPal was conducted with one-on-one use *i.e.*, a student sitting at a computer, so it was uncertain how differently students would interact with the system as part of social groups or of family meetings.

PRODUCING A VIRTUAL CONVERSATION

This section lays out ICT MentorPal's current operations data flow. The user inputs text or voice and the rest of the flow, as shown in Figure 1, is a succession of parsers and filters to select and cue up the most appropriate answer. The creating a large data base of short clip files is a somewhat daunting task, described in earlier papers. (Nye *et al.,* 2017) methodology for asking questions, turning questions into transcripts, cleaning transcripts, feeding these into dialog models, and improving these models by classifying new paraphrases/aliases for questions. Assembling an interactive virtual mentor from previously recorded videos takes up to approximately 20 hours of video taping to provide a broad enough coverage of useable responses, with approximately 60 hours of additional time for production processing: about 40 hours for preprocessing transcripts and 20 hours for post-processing dialog models. Because a long-term goal of this project is to enable the easy recording any person as a mentor, without

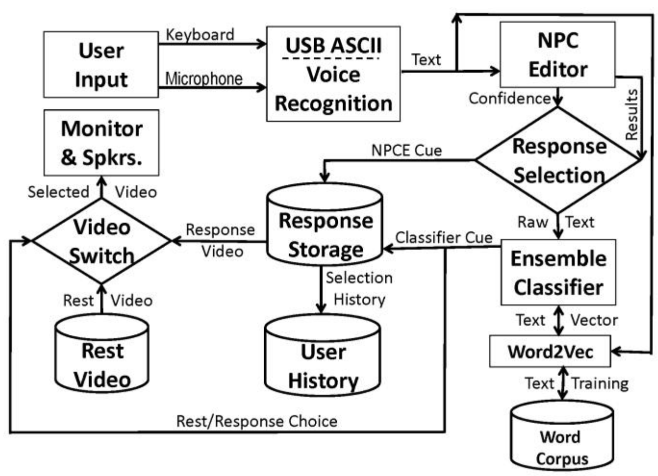


Figure 1.MentorPal Conceptual Flowchart

researcher intervention, improvements to the current pipeline are noted when discussing each step of the process.

Drafting the Questions and Recording the Mentors

The question set that was created (Nye, et al., 2017) consists of four types of questions: 1) preset questions that are asked for all mentors (˜70% of questions), 2) mentor-specific questions related to their career or life experiences (˜10% of questions), 3) follow-up questions identified during the interview that seem relevant to responses given by a specific mentor (˜10% of questions), and 4) mentor-specific common questions observed from chat logs by students with a mentor (˜10% of questions). This process is outlined in Figure 2. The preset questions span a variety of topics detailed in Table 1 (below). Each mentor was recorded across recorded across at least five sessions, each about three hours long, to cover the question set. These sessions have a mix of these topics, with a follow up section for each session. After the initial sessions, re-recording sessions are held to record new answers to bad video clips or to record answers to common questions that were not recorded, *i.e.*, the fourth type of question. Each follow-up session typically lasts for about 45 minutes to an hour. These follow-ups are important because they assist with creating a more natural conversational flow, by enabling the mentor to answer questions that are likely to be asked after a student has heard a particular response.

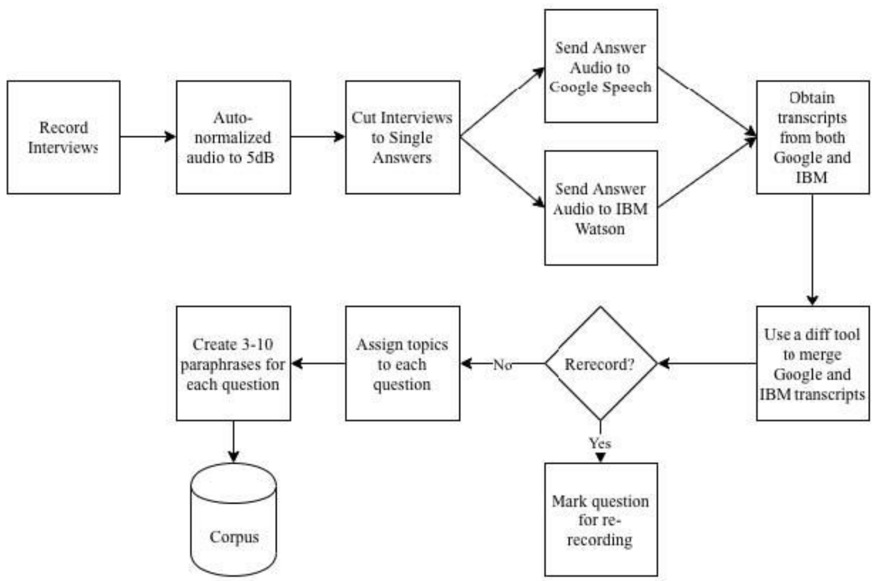


Figure 2. Examples of Questions by Topic

Recording currently involves calibration with help from project staff in order to ensure that the recordings are consistent and the responses are appropriate to chunk into useful videos. Standardization includes ensuring that the mentor maintains a resting position 3 seconds before and after the question is answered. The resting position should be in the same spot throughout the recordings, to avoid "jumping" where a mentor's next statement has a totally different head or body position. In addition to this, the mentor is recommended to try to use emotion, smiles, stories, and hand motion to provide engagement. So far, each recording has been standardized by using the same low-cost equipment setup, consisting of professional quality webcam, microphone, and backdrop, which is shipped to mentors who are located remotely. After recording, either the mentor or the research staff uploaded the session files to cloud storage for post-processing. Overall, this approach has worked effectively, with the quality of recordings and transitions significantly higher than was initially anticipated. Sound quality has occasionally suffered, however: mentors have forgotten to activate the provided microphone, using a laptop internal microphone instead or the use of the professional microphone incorrectly, such as facing the wrong way for the very directional microphone.

An area of improvement for this part of the methodology would be to try and move this to a standalone process where the mentor could record follow up questions themselves. This is challenging, because it is hard for a mentor to predict what follow-up questions a student might ask. An automated mechanism to help detect likely follow-up questions, *e.g*., detecting uncommon proper nouns, might help a mentor self-record more effectively.

Preprocessing for Transcripts

At this juncture, the individual questions are rendered into transcriptions and video chunks. To prepare the sessions to run effectively through the preprocessing Python script, the start times and end times are manually annotated to for each question answer. The Python script then takes each MP4 file and audio normalizes it with FFMPEG-normalize, chunks it up according to the timesheet document, converts the chunks into sound Waveform Audio Files (.wav) and then makes a call to IBM Watsons Speech-To-Text API. The call to the Speech-To-Text API takes around one hour to complete for 18 hours of footage. However, a major bottleneck for this process is the manual cleanup of the transcription. The number of errors is significant enough to warrant cleanup, at least if learners are meant to read the transcripts, which is important for deaf learners, users in high-noise areas, or users without sound capability. Moreover, the errors tend to be systematic: they happen most frequently for domain-specific terms about a career, for words that a mentor says with a regional accent, or where sound quality was poor. As a result, transcripts are manually edited by listening to the audio from the recordings. This process takes twice as long as recording, over 40 hours to edit 20 hours of video.

This segment of the procedure has the potential for the most improvement, given the large cost of personnel time. Currently, research is being done if other Automatic Speech Recognition services like Google ASR would be an alternative or complement to IBM Watson, which is an ensemble model approach. Initial research on this has not yet produced positive results, but is still ongoing. In addition to changing the ASR, improvements on the translation cleanups could make better use of the translation confidence to define a threshold to detect answers that need to be manually edited. In regards to improving the editing workflow, text "diff tools", flagging differences, that could identify any discrepancy between two different transcripts. Alternatively, audio-aligned text editors could be used to more rapidly parse transcripts, as opposed to searching for the part of the video that matches a poor transcription.

Processing the Recorded Clips for Engaging Dialogues

An ensemble model classifier chooses an answer from the two classifiers. Currently, it sends it to both the NPCEditor and a python classifier based on a neural network to get two alternate answers. Of these, the NPCEditor answer is used if it has high confidence (since it is faster) while neural network classifier is used otherwise. Further description of these processes is contained in (Nye, *et al.*, 2017). This section describes how the data for these models is created. As shown in Figure 2, question and answer text pairs are used to produce neural-network classifier data, NPCEditor data and metadata (Leuski & Traum, 2011, and Nye, *et al.*, 2017). The initial paraphrase sets are generated manually based on based off of other ways that we think other people would ask the question. We also manually tag each question with topics from a bank of 40 predefined topics that cover all broad categories of questions. Topics are used in order to generate random questions by topic and are also added to the neural network classifier data as a sparse vector for each question. It then dumps all this data into files for use by the neural network classifier. The NPCEditor reads an analogous file generated from the same data, which contains only the paraphrases/aliases of the same questions.

The procedure for creating such a conversational VH mentor involves a number of steps, in addition to the development and tuning of the software code that makes the program function. Literally volumes have been written about how to system engineer large code programs, running all the way from general, almost philosophical approaches, *e.g*. Professor Fred Brooks book (Brooks, 1995) to more detailed and didactic tomes (Pressman & Maxim, 2005). However, for the more prosaic support functions, no such tomes are familiar to the simulation computer scientists that would assist in organizing their activities. Project personnel, basing their approach on lessons learned from previous ICT efforts, knew they needed a process for the production of the videos.

Research is ongoing to continue improving these processes. A particular issue with the classifier was the desire to enable offline mode on tablet devices. Unfortunately, the word vector models employed are typically too large to work on device in memory: the Google News Word2Vec model is 3.5 GB, for example (Mikolov, *et al*., 2013). To address this, the ensemble makes use of a systematically pruned version of Google's Word2Vec model described more fully in a workshop paper (Kaimakis, Davis, Beck & Nye, 2018).

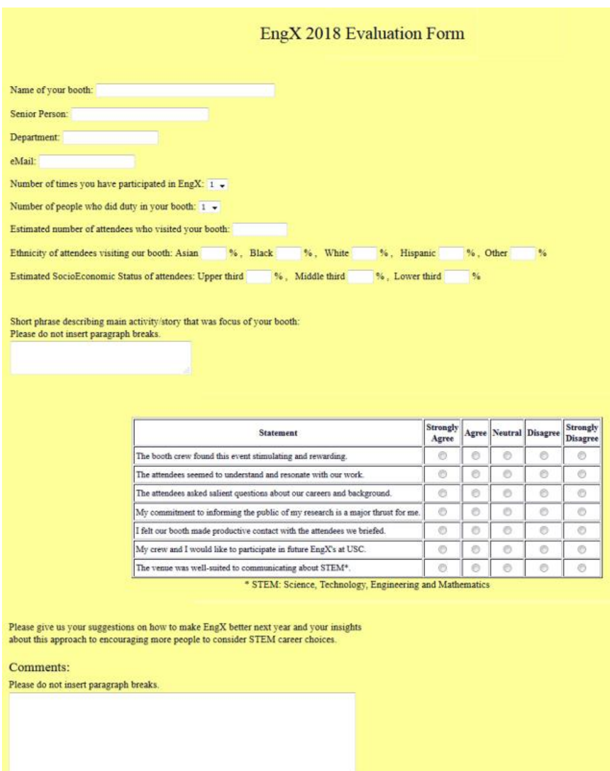
EXAMPLE IN PRACTICE: ENGX STEM FAIR

The present version of the MentorPal system was tested with two mentors at an event called the EngX 2018 STEM Fair, see Figure 3. The fair is part of a series of STEM events conducted over the last four years, led by the Armani Lab at the University of Southern California (USC). This year's fair was attracted a broad range of students aged from eleven to eighteen years old. It was a full-day fair, running from 8 AM until 4 PM and was held on a Saturday in March on an open quadrangle on the USC campus and attracted several hundreds of students, parents and teachers. EngX contained about 20 booths and the majority of exhibits were not computer-based, but were instead physical: water-powered rocket cars, pipe mazes with miniature robots, mirror-based holograms, and similar demonstrations. Some booths used posters from conference poster session, which seemed to be less enticing than other approaches.

Figure 3 EngX on a USC Quad

Students arrived in varying-sized of groups to the various booths and some came alone, most in small groups and some were closely attended by one or more adults/parents. They were observed to be of varying socio-economic status, but tended to be above the mean of the spectra. Like many grade school students they seemed more attracted to interesting new experiences than sober contemplation on the career choices they were about to make. As such, in addition to using the system, many visitors were curious about the goals of the work and how it related to the students' own interests in research or technical work.

The EngX venue seemed to be well-suited to communicating about STEM, both in person and via MentorPal. Both students and parents evidenced and articulated the need for the information being proffered.



EngX 2018 Evaluation Form

Name of your booth:

Senior Person:

Department:

eMail:

Number of times you have participated in EngX: 1

Number of people who did duty in your booth: 1

Estimated number of attendees who visited your booth:

Ethnicity of attendees visiting our booth: Asian %, Black %, White %, Hispanic %, Other %

Estimated SocioEconomic Status of attendees: Upper third %, Middle third %, Lower third %

Short phrase describing main activity/story that was focus of your booth:
Please do not insert paragraph breaks.

Statement	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The booth crew found this event stimulating and rewarding.	○	○	○	○	○
The attendees seemed to understand and resonate with our work.	○	○	○	○	○
The attendees asked salient questions about our careers and background.	○	○	○	○	○
My commitment to informing the public of my research is a major thrust for me.	○	○	○	○	○
I felt our booth made productive contact with the attendees we briefed.	○	○	○	○	○
My crew and I would like to participate in future EngX's at USC.	○	○	○	○	○
The venue was well-suited to communicating about STEM*.	○	○	○	○	○

* STEM: Science, Technology, Engineering and Mathematics

Please give us your suggestions on how to make EngX better next year and your insights about this approach to encouraging more people to consider STEM career choices.

Comments:
Please do not insert paragraph breaks.

Figure 4. Questionnaire for Booth Crews

There was no known attempt to formally characterize the attendees, but the ICT booth crew observed a number of characteristics that seemed to bear on the issues at hand. Parents were particularly active in pursuing issues of concern to them, even to the detriment of the time their students had with the ICT professionals. They found it advantageous to have one member take the parents' question, while another took the students in hand and both talked to them and assisted their use of the MentorPal program.

To try to get a better view of the range of experiences of entire EngX fair, the ICT booth crew created an informal *post hoc* questionnaire '(see Figure 4) and data manipulation program using HTML and PHP to get some sense of what the other booth crews had observed, and put it up on line. Personnel from a small group of the booth crews have responded. Rather than give a false impression of precision, perhaps the most conservative approach would be to merely report the generalities supported by the responses received to-date, with an assurance that all other responses will be characterized in the oral briefing in November.

The attendees were significantly skewed toward the ethnicities of Asian and European, despite the fact that the immediately contiguous residential areas are heavily populated by traditionally under-represented minority groups: Black neighborhoods to the south and Hispanic neighborhoods to the west. Further, all the respondents estimated the Socio-Economic Status (SES) of the attendees as being from the top two thirds of the population. This data is especially suspect, as the SES was estimated by attire and manifest erudition of the parents and the students.

As to the five item Likert Scale survey, they responded to six statements, which are set forth below:

1. The booth crew found this event stimulating and rewarding.
2. The attendees seemed to understand and resonate with our work.
3. The attendees asked salient questions about our careers and background. (this got the lowest Likert score)
4. My commitment to informing the public of my research is a major thrust for me.
5. I felt our booth made productive contact with the attendees we briefed.
6. My crew and I would like to participate in future EngX's at USC.

The responses were all "firewalled" to the right (Neutral, Agree, Strongly Agree), with virtually none receiving any "Strongly Disagree" or "Disagree Marks." Our hope of correlating some of the demographic data with the Likert responses has, so far, been frustrated by the low response rate of the already fairly limited number of all of the booth crew participants. Nonetheless, the responses do convey data from which the reader may draw some insights. Of particular note was the nearly universal commitment to participate in next years' EngX and the high marks given to MentorPal by users and by the parents.

The "traffic" through the MentorPal booth was high, with approximately 50% of passing groups stopping for a significant span of time. This was likely partly due to being the exhibit that was the most clearly AI-relevant, as well as due to the use of a large banner poster (see images of the EngX fair in Figure 5 below). Three distinct types of interaction emerged when exhibiting MentorPal in this context: 1) group briefings on work with large screen display, made up of typically two or more students accompanied by an adult or two, 2) individual interactions with MentorPal on individual tablet computers, who were almost always students, and 3) individual discussions about the work and technical careers, more often initiated by parents than students.

Figure 5. EngX Fair MentorPal Setup

Perhaps one in ten of the small groups who came to the MentorPal booth went on to other booths, and then returned to ask more questions or get back on the program. As he booth crew were almost constantly engaged counseling, they had little time to record events like these. A lesson learned is that one should "over-staff" a booth, to ensure data collection and photography. Figure 5 is made up of two images that were taken, showing the banner used to attract students and their parents as well as a typical group being briefed by the MentorPal project’s Principal Investigator.

This exhibit reaffirmed findings from single-user pilot testing where students often were unable to think of any useful questions for MentorPal, *e.g*., freezing or asking questions unrelated to careers). This was particularly relevant for students accompanied by parents. However, students were more productive when they used a menu that allowed generating questions from topic categories the team thought the students should be asking. In this fallback mode, rather than entering free-text questions, they clicked on the topics to see suggested questions to ask. While the menu-based experience resulted in a more traditional question and answer, dialog, there are advantages to this mode: students may learn from reading the question suggestions and it sidesteps classification errors. By comparison, MentorPal appeared to respond to free text questions with reasonable answers about two-thirds of the time, while in other cases answered inaccurately or had no answer. This is apparent accuracy is higher than the actual accuracy: leave-one-out testing found only about 50% for hold-out paraphrases matched their equivalent question (Nye, *et al*., 2017). However, the performance of the classifier appeared higher in practice, since many “misses” still result in MentorPal replies that are reasonable to the user. Individual misses that were noted by learners did not seem to break engagement, but multiple in a row appeared to reduce interest. Overall, while classification accuracy might be beneficial, the larger issue appears to be that students struggle to ask productive questions.

For students using the system individually, they were content to ask questions from MentorPal for fairly long periods of time, approaching fifteen minutes. However, it should be noted that these students were self-selecting. As the intent of the project was to create a conversational easiness with the student that would encourage the mentor-mentee relationship, one of the areas in which the team was most interested was the students' view of their dialogue with the mentor. Since this usage was as an exhibit, rather than as a research study, no formal survey was done of their views. However, to help gather feedback about how to improve the system and know visitor reactions, a straw poll was offered to students and adults who spent substantial time interacting with the mentors, either individually or as part of a group. The results of this straw poll are summarized in Table 1. To note, the age shown was the median age. Adults were more likely to complete the straw poll, while students were more likely to leave without filling out a poll. Overall, less than half of visitors who stopped for at least a minute took the straw poll. 44 completed it, implying a lower bound of 80 visitors and likely more than 100. Reactions were positive overall for those who completed the poll, with 77% agreeing that the answers were conversational and 84% enjoying the experience. With that said, this sample is inherently biased: those who used the system for less time would likely also skip the poll, so this sample likely represents more motivated users.

Responses from Users

Md=15.6 N = 44

Age: $\bar{X}$ = 18.6 Gender: M: 27 F: 17 College Major/Career Choice: Mostly STEM

Please Circle the number that best describes how you feel.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. I think or worry about my future life almost every day.	4	5	22	6	7
2. Having easy access to better career advice would be useful.	0	0	17	18	9
3. I liked the on-line mentors and learned from them.	0	0	7	26	11
4. The mentor's answers were on-target and useful.	0	0	18	16	10
5. The responses seemed "conversational" and real.	0	0	10	15	19

Table 1 Straw Poll Results

Among the new insights emerging from the MentorPal participation in EngX was in terms of its potential impact on the prospective users. That was the interaction between the parents and the students. On the one hand, the booth crew reported that the parents were much more focused on the career choice facing their students. This seemed to drive productive questions to and answers from MentorPal on career selection and planning. However, this also caused the students to become more withdrawn and incommunicative; often the parents talked and the students just stood. While this is not hard evidence of such a trend, all of the booth crew reported the same impression. By comparison, one-on-one interactions between a student and the virtual mentor tended to separate the parent as an intermediary. The tradeoff of this case was that students varied significantly in how focused their questions were, *e.g.*, more off-topic or asking about biographical details such as family members, rather than career focused.

Since students likely benefit more from the one-on-one interactions, *e.g*., can ask their own questions, research is now looking into dialog techniques that help steer them toward more productive questions. Different strategies are being considered, which include the importance-weighting and politician responses. Importance weighting would consider the research team's assessment of the quality of the answer during the selection process, *e.g.*, rather than answer the "best match" from the classifier, instead use a weighted mixture that considers the content value. This should be particularly useful for vague questions, which are common for free text questions, *e.g*., "What do you like?".

RESEARCH DIRECTIONS: POTENTIAL SOLUTIONS TO LIMITATIONS

Based on the results from testing with students so far, a number of possible approaches have been identified to improve the experience and information shared during a brief virtual mentoring session using this technology.

Expanding the Mentor Experience to Multiple Mentors

An issue in presenting virtual mentors is that learners may not want to be constrained in to talking to only one virtual mentor, particularly if they feel they did not find that mentor interesting. A series of one-on-one chats also makes it hard to see contrasting experiences. One approach to solve these issues is to consolidate mentors into a panel, where students can hear answers from several agents. Each agent responds to the question and the most relevant and accurate answer is displayed. Students can then listen to the other mentors' responses to get a wider variety of experiences and opinions. Students may also lock a mentor to engage in one-on-one dialog with that agent.

The main benefit of this approach is that students hear from professionals spanning multiple STEM careers and see how those fields differ from each other. Students who lack exposure to STEM careers may not know which fields they can explore, and this approach will give them an overview of several different career paths. (see Figure 6) They can then dive into a specific career or mentor that interests them. Having multiple mentors should also reduce misses and vague answers, as mentors with irrelevant or low-scoring answers can be ignored in lieu of those with better responses (*i.e.,* mentors without a good answer are grayed out).

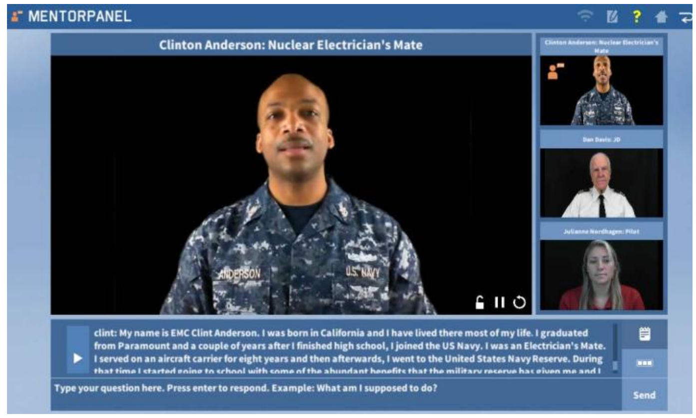


Figure 6. MentorPanel Prototype

A limitation of the panel dialog is the inability of mentors to interact with one another. The mentors are recorded separately so they cannot react to or discuss with other mentors. This makes conversations comparably one-dimensional and less engaging than they could be with a live panel. While some consideration has been made about how to approach the issue of non-interacting panelists, it would be fairly challenging to do using video-recorded mentors. However, virtual agent mentors (or a virtual agent moderator) could potentially be used to prompt mentor interactions such as opposing viewpoints.

Assisting the User with Salient Issues

While the first mentors to be recorded addressed STEM careers with answers that fairly were cogent and compelling, users' ability to ask good questions was a primary limiting factor. This suggests that there is a pedagogical issue which must be addressed. For example, when piloting this research with the Next Generation Leaders mentoring program, their instructors preceded the interaction with a brief exercise where students discussed their biggest role models and the kinds of questions that they would want to ask them. After the interaction, they then reinforced lessons-learned by having a discussion about what was most interesting and about what parts of careers the virtual mentor helped them think about. This was a novel interaction, and the research team is interested in seeing how different teachers might integrate this kind of technology into larger lesson plans for their students.

This strategy might also be possible for virtual mentors to follow. This would require the virtual mentor to pose significant questions and respond meaningfully to those responses. This situation brings us a whole new area of conceptualization and communication. At least for small question progressions, this might be accomplished even with video-based mentors. This could be done with limited recording time with a small set of standardized questions like "Who are mentors in your life?", "What are you worried about for your career?", or "What parts of a career do you think are most important?" While processing these would require careful attention to the scope, question asking has been done using this technology if kept well-bounded.

CONCLUSIONS

Both technical career events like EngX and on-line resources demonstrate a good potential for overcoming the nationally serious issue: the declining availability of competent technical personnel. The career fair approach is personal, engaging, and illuminating, but there is still the issue of getting under-represented groups to attend. The on-line method is available to low SES students and those who are geographically remote. It is clear that there are many improvements on the horizon, but the system already shows promise as a mentor conduit: allowing the most engaging set of mentors to communicate with a broader population of students.

The ICT experience at the STEM fair indicated that students do engage effectively with this sort of virtual human conversation system in a one-on-one environment, but with they do need a suitable introduction to the goals and interactions with the system. While the amount of time for such interactions is only moderate, say up to 15 minutes, even for a small booth with five tablet computers students spent many cumulative hours speaking with the mentors, which is many more than the booth staff would have been able to accomplish alone. Given the limited availability of mentors and the desirability of instant access to a mentor on-demand, this particularly indicates that an on-line delivery system for mentors would be ideal, as the current version runs on-device.

There is a putative benefit of this model, which is the opportunity to capture particularly engaging and compelling human mentors, even those who might no longer be available. For example, there is the case of Jaime Escalante, the renowned high school teacher of calculus. There are those who argue his skills were literally unique ones based on charisma and personal characteristics which cannot be readily taught or transferred to other teachers or role models (Jesness, 2002). Such mentors could communicate concepts for classroom instructor who would not have the same skill set.

A final concern is the adoption and support of this technical advancement by the educational establishment and counseling community. The authors are aware of many programs like Project Seed (Phillips & Ebrahimi, 1993) which showed promise but were not adopted by the education leadership. It could be argued that the most effective implementation of this approach can only be realized with the support of the teaching community and existing networks for STEM mentors (Smith & Anderson, 2014). Such institutions are as a source for recruiting and recording mentors and as venues for encouraging interaction with virtual mentors. They would be the best suited to know who represent effective mentors for their student groups, where an ideal case might be to capture a highly-impactful real-life local mentor who is expected to become unavailable, *e.g.*, move away, new family responsibilities. Toward that end, there is a strong interest in identifying how to help teachers and similar leaders record and refine mentors.

To sum up, the need for attracting the best candidates into technical work is a critical national trend. The authors and others continue to recognize huge gaps between the assumed familiarity of students with technical careers and personnel, which can lead to unsustainable career choices. Both technical career fairs and virtual conversational mentors have a place in attacking these issues. The ones most in need are the lower SES children and those located away from major centers of technical and academic activity, but even in technical homes, the authors find little effective communication about what a technical person's workday looks like.

These observation lead to many open and desirable future research paths. Work in web based systems, creative conversational initiatives, and body-language/facial expression monitoring are offered as major priorities. Also on the docket of desired work is the design and fielding of mentor creation by computer users with little or no knowledge of programming or video production.

ACKNOWLEDGEMENTS

More than anyone else, the authors wish to express their appreciation for the MentorPal Principal Investigator, Dr. Benjamin Nye. His support and guidance is the foundation on which all this work was accomplished. MentorPal is supported by grants from the Office of Naval Research from the ONR STEM Program: Mentor PAL: Growing STEM Pipelines with Personalized Dialogs with Virtual STEM Professionals, N00014-16-1-2820. The statements and views expressed in this paper are the views of the authors alone and do not represent the views of the Office of Naval Research or the US Government. Additionally, we would like to acknowledge the leadership of Dr. Andrea Armani and the hospitality of the Armani Lab at the University of Southern California who organized the EngX STEM fair where MentorPal was exhibited.

REFERENCES

BLS, (2014), STEM 101: Intro to tomorrow's jobs, U.S. Bureau of Labor Statistics, Washington D.C., retrieved on 05 April 2018 from: https://www.bls.gov/careeroutlook/2014/spring/art01.pdf

Brooks, F. (1995), *The Mythical Man - Month. Anniversary Edition*, Boston, MA: Addison-Wesley Professional

Chowdhury, G. (2003). Natural Language Processing. *Annual Review of Information Science and Technology*, 37(1), 51-89.

Collobert, R., & Weston, J. (2008, July). A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning." In Proceedings of the *25th International Conference on Machine Learning* (pp. 160-167).

Dimitrova, V., and Brna, P. (2016): From Interactive Open Learner Modeling to Intelligent Mentoring: Style-OLM and Beyond. *International Journal of Artificial Intelligence in Education* 26(1), 332–349

Jesness, J. (2002): Stand and Deliver Revisited. *Reason* 34(3), 34–39

Kaimakis, N., Davis, D., Breck, S., Nye, B. (2018, Pending): Domain-specific Reduction of Language Model Databases: Overcoming Chatbot Implementation Obstacles. In: *Proceedings of the ModSim World Conference*.

Lederman, N. G. (1992). Students' and Teachers' Conceptions of the Nature of Science: A review of the research. *Journal of Research in Science Teaching*, 29(4), 331-359.

Leuski, A., and Traum, D. (2011): NPCEditor: Creating Virtual Human Dialogue using Information Retrieval Techniques. *AI Magazine* 32(2), 42–56

Maoldomhnaigh, M. O., & Hunt, A. (1988). Some factors affecting the image of the scientist drawn by older primary school pupils. Research in Science & Technological Education, 6(2), 159-166.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J. (2013): Distributed representations of Words and Phrases and their Compositionality. In: *Advances in Neural Information Processing System*s. pp. 3111–3119

Nye, B., Swartout, W., Campbell, J., Krishnamachari, M., Kaimakis, N., Davis, D. (2017): MentorPal: Interactive Virtual Mentors Based on Real-life STEM Professionals. In: *Proceedings of the Interservice/Industry Simulation, Training and Education Conference* (I/ITSEC).

Phillips, S., & Ebrahimi, H. (1993). Equation for success: Project SEED. Reaching all students with mathematics, 244.

Pressman, R. S. & Maxim, B. (2005). Software Engineering: a Practitioner's Approach. Palgrave Macmillan, Basingstoke, UK,

Rizzo, A., Lange, B., Buckwalter, J.G., Forbell, E., Kim, J., Sagae, K., Williams, J., Difede, J., Rothbaum, B.O., Reger, G., et al. (2011): Simcoach: an Intelligent Virtual Human System for Providing Healthcare Information and Support. *International Journal on Disability and Human Development* 10(4), 277–281

Smith, A., Anderson, J. (2014): *AI, Robotics, and the Future of Jobs*. Pew Research Center 6

Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., Williams, J., Leuski, A., Narayanan, S., *et al* ,Piepol, D.,. (2010): Ada and grace: Toward Realistic and Engaging Virtual Museum Guides. In: *International Conference on Intelligent Virtual Agents*. pp. 286–300. Springer

Traum, D., Jones, A., Hays, K., Maio, H., Alexander, O., Artstein, R., Debevec, P., Gainer, A., Georgila, K., *et al.* Haase, K., (2015): New Dimensions in Testimony: Digitally Preserving a Holocaust Survivors Interactive Storytelling. In: *International Conference on Interactive Digital Storytelling*. pp. 269–281. Springer

Wang, J., Li, H., Cai, Z., Keshtkar, F., Graesser, A., Shaffer, D.W. (2013): Automentor: Artificial Intelligent Mentor in Educational Game. In: *International Conference on Artificial Intelligence in Education*. pp. 940–941. Springer