Kevin Zyskowski

Professor Savage

5170 HCI for AI

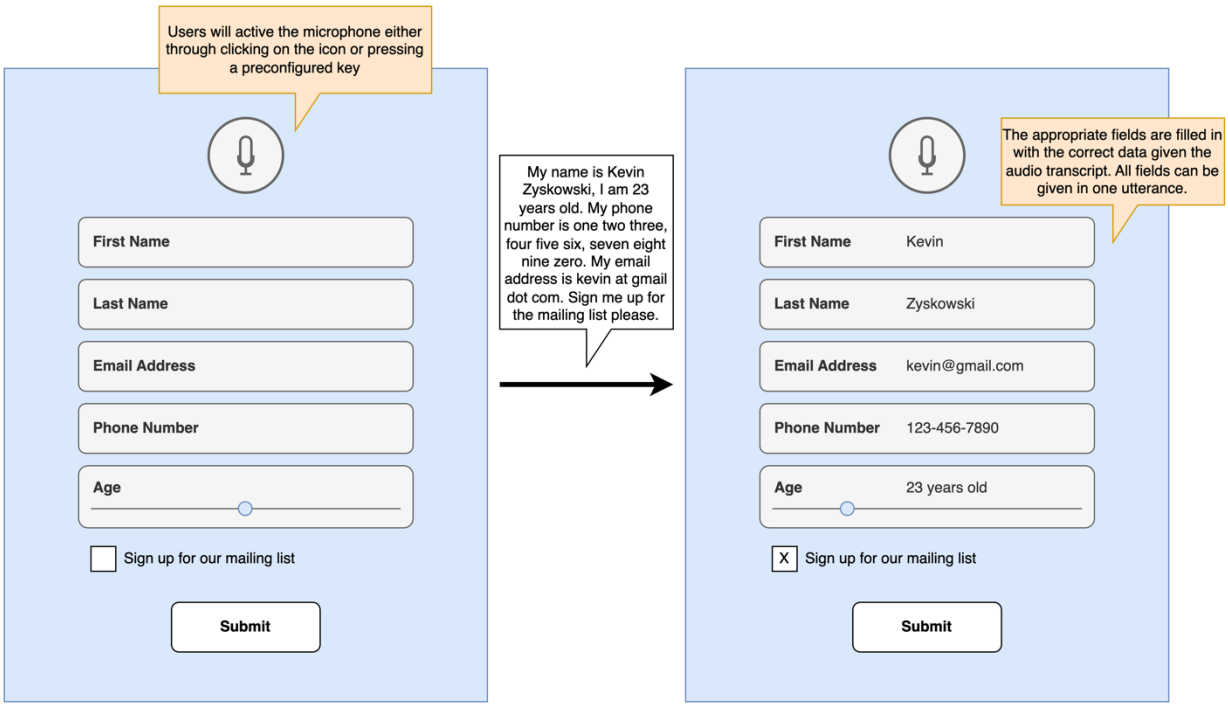
23 November 2024

Web Forms Through Large Language Model Dictation Systems

For this project, I've designed a proof-of-concept AI system aimed at improving dictation software for individuals suffering from motor impairment disabilities such as quadriplegia. Dictation software allows individuals to audibly voice commands such as "tab" and "enter" to mimic keyboard actions and navigate computer systems (Level Access How do people with disabilities access the web?). This is a time-tested method that allows individuals with motor impairments to complete various tasks such as online shopping, viewing/sending emails, and more. When writing documents, speech-to-text systems even allow users to speak natural language and automatically commit it to text. However, for more complex inputs such as web forms, individuals must navigate individual fields using dictation commands before entering information. I believe that this is an unintuitive system for filling out web forms.

I am targeting web forms to demonstrate my minimum viable product and illustrate how Large Language Models (LLM) are incredibly powerful at bridging the gap between natural language and machine interfaces. Prior research has been conducted into form completion software using speech recognition but requires dictation of field names as well as their values to help systems navigate to appropriate input fields (Issar, 1997). Modern LLMs can potentially overcome this barrier.

In this prototype, the user first activates voice recognition and then records audio capturing relevant form information in natural language. After recording, the LLM parses the audio and creates a text transcript, and then the LLM parses the text transcript and outputs the extracted fields in JSON format, making it trivial to fill in fields programmatically. We can see an example user flow in the wireframe below:



I was inspired by the presentation given by Professor Dakuo Wang during class in the past week, specifically by the Alexa demo in which multiple inputs are given in one utterance. This is a more intuitive approach to filling out forms audibly because natural language is not limited to communicating information one field at a time. For example, I can provide my full name as Kevin Zyskowski instead of separately providing my first name as Kevin and my last name as Zyskowski. LLMs excel at extracting these kinds of relationships from inputs.

It is important to highlight potential harms that can be made to the user. One issue that arises concerning security and privacy is the fact that 3rd party models offered by companies such as OpenAI provide their LLMs over an internet API. If potentially sensitive forms are filled out by individuals (including fields such as SSN or account passwords), then it is important that audio/text files sent to OpenAI are protected from attackers. One potential mitigation strategy is to opt for local models for transcribing audio and extracting form information. In the future, once more powerful models become local-first, this approach will be more tractable.

In the future, I believe a design such as this would be very beneficial if developed into a browser extension/addon that can automatically detect <form> and <input> tags and work out-of-the-box. In the basic prototype I programmed, the form is specially designed to accommodate LLM-augmented completion, but I explored various types of input fields including text, number sliders, and checkboxes. Immediate next steps would include experimenting with various form types, and designing a generalizable framework that can be applied to forms outside my example web application.

I have provided a short video demoing the speech-to-text form completion functionality to accompany this document.

References

- Level Access. “How Do People with Disabilities Access the Web?” *Level Access*, 5 Nov. 2024, www.levelaccess.com/blog/web-access-people-with-disabilities/.
- Issar, Sunil. "A speech interface for forms on WWW." Fifth European Conference on Speech Communication and Technology. 1997.