Evaluating Conversational AI and Specifications: SpecAI

Sergio Riccardi & Spyridon Ampanavos

Perkins&Will

Introduction

The objective was to test the hypothesis that incorporating Conversational AI into specifications could lead to quicker responses, improve the interaction between project teams and project manuals, and aid less-experienced designers in navigating project manuals.

→

Road map for our innovation incubator

Research

Interviews

Study Build

Study launch

Results
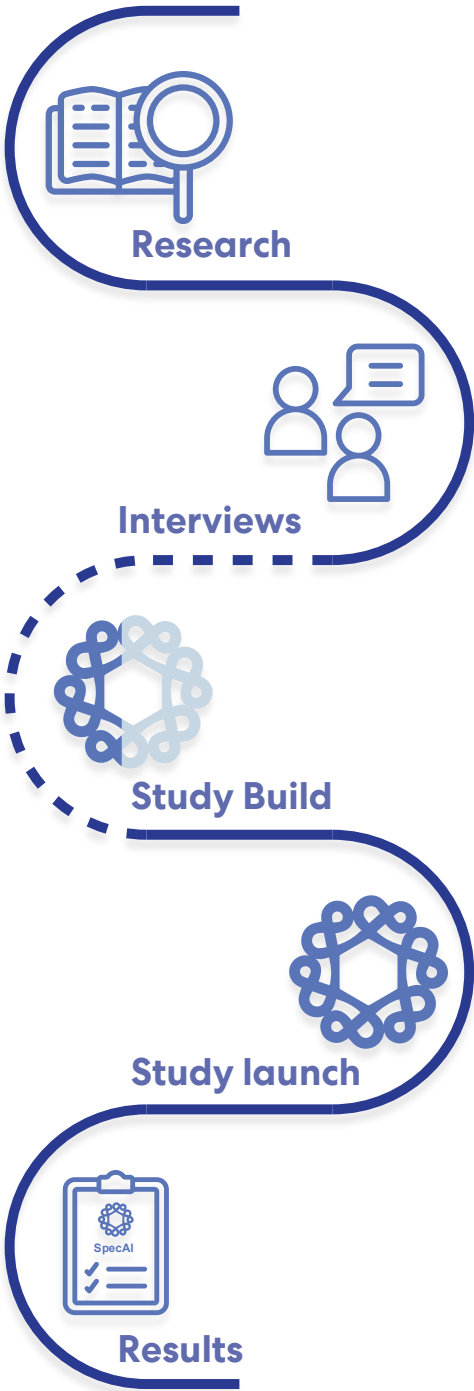
Background

What's the problem / what are the characteristics

Large files, search and review may be difficult/tiring

Spec writing: previous work not easy to reference

Steep learning curve for junior designers.

**The technology**

ChatGPT

Google Bard

Bing AI

Implications

Here is an overview of our initial research findings concerning applications of conversational AI that bear relevance to the concept of SpecAI and our goals with this emerging technology. These findings can be adapted to align with the application we intend to develop within the ACE Profession. Many of these findings are loosely connected to the concept of SpecAI.
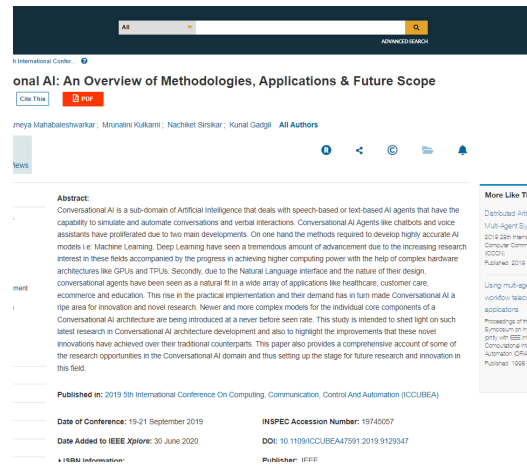
Conversational AI: An Overview of Methodologies, Applications & Future Scope

Date: 2019-09
Authors: Pradnya Kulkarni, Ameya Mahabaleshwarkar, Mrunalini Kulkarni, Nachiket Sirsikar

Conversational artificial intelligence in the AEC industry: A review of present status, challenges and opportunities

Date: 2023-01
Authors: Abdullahi B. Saka, Lukumon O. Oyedele, Lukman A. Akanbi, Sikiru A. Ganiyu, Daniel W. M. Chan

"So what if ChatGPT wrote it?

Date: 2023-08
Authors: Yogesh K. Dwivedi, Nir Kshetri, Laurie Hughes, Emma Louise Slade, Anand Jeyaraj, Arpan Kumar Kar

The Morphological Echo of Architects. Concept for a Conversational Artificial Intelligence to Support Architects during the Early Design Stages
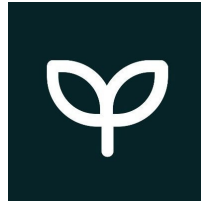
Date: 2022
Authors: Jessica Bielski

In Practice:
Using your Documents with LLMs

AskYourPDF   **ChatPDF**   **PDF.ai**   **Filegpt.app**   **AIPDF.app**

The majority of these applications operate by utilizing OpenAI's ChatGPT API integrated into a bespoke user interface, with the inclusion of custom instructions to ensure that responses remain strictly within the context of the provided information. During our review of these applications, OpenAI introduced new functionalities, such as Plug-ins and Custom Instructions, which enabled us to harness ChatGPT in a more controlled fashion. Consequently, we chose to construct the user study around these advancements, and we were able to employ AIPDF.APP's GPT plug-in to execute the study.

# (How) Can text GenAI be used to improve our spec processes?

Method

Methods

Interviews:

Conduct interviews with project team members, the General Counsel at Perkins&Will, and specification writers to ascertain their level of interest in AI, identify constraints within their current work processes, and pinpoint any reservations or concerns related to the integration of AI into their workflows.
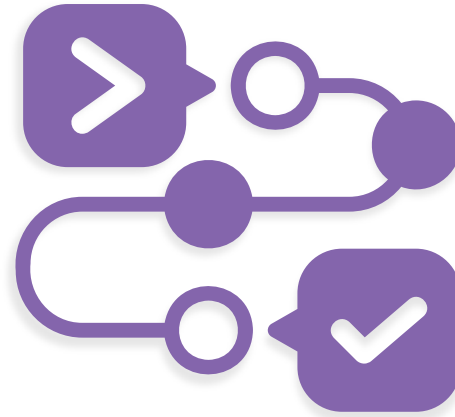
Developed a tool/workflow.

Examine the existing consumer market tools designed for AI-assisted PDF review and develop a workflow based on insights gathered from our interviews. This workflow should seamlessly integrate with our current Submittal Review process.
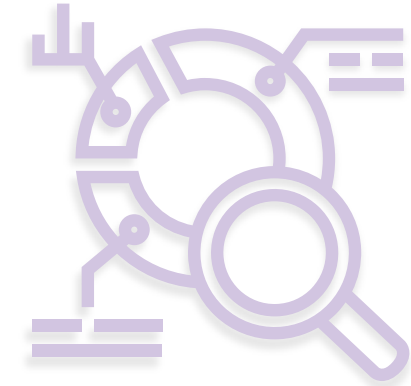
Evaluate with user study.

Following the user study, analyze the results to discern the most effective practices for utilizing AI tools in providing Construction Contract Administration (CCA) assistance.

Interviews:

Prior to commencing the development of our user study, a series of interviews were conducted with various roles within a typical project team. These roles included Junior Designers, Project Designers, Project Architects, and Project Managers. While we acknowledge that there are additional roles involved in the design process, our primary focus was on those who engage with specifications during the Design Development (DD), Construction Documents (CDs), and Construction Contract Administration (CCA) phases. Additionally, interviews were conducted with Specification Writers to gain insights into current technologies and their specific needs.

Furthermore, we conducted interviews with Perkins&Will's General Counsel to address legal concerns and ensure full compliance in the construction of our user study. For each role, we formulated a questionnaire, and the following are some illustrative examples of these questionnaires:

Project Designer / Designer i, ii, iii   **Project Architect**   **Project Manager**   **General Council**

Interviews:

Harrison Maki
Designer I

> "Ideally the AI would show you where to look for information and then you can cross reference it yourself"
>
> - Harrison Maki

Jeffrey Brussel
Senior Project Architect

> "I mean, even if you wanted to be against it, it's coming. its like adapt or die"
>
> - Jeffrey Brussel

Marko Goodwin
Specification Writer

> "There's like a whole art of just being able to read specifications, like having literacy, AI could help bridge this gap"
>
> - Marko Goodwin

Joelle Jefcoat
Deputy General Counsel

> "No one has yet brought something to us where they're actually using conversational AI for a business purpose."
>
> - Joelle Jefcoat

The tool/workflow

Chat GPT
+
Prompt Engineering: Custom instructions/ conditioning
+
Specifications Document
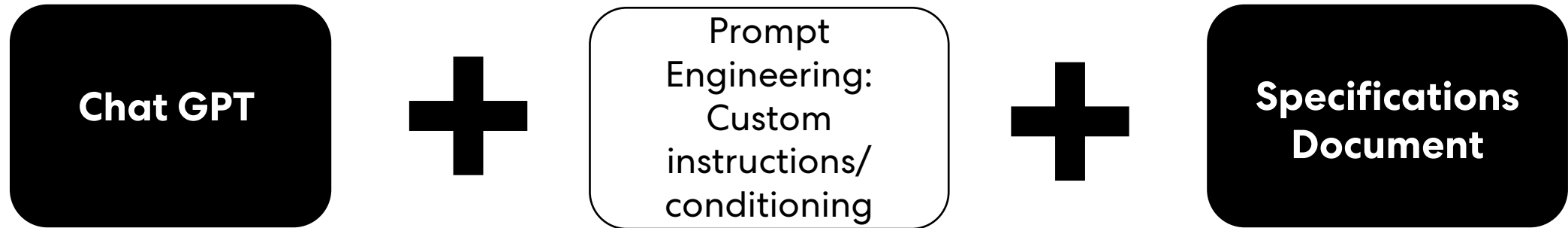
How AI PDF works

How to Setup:
User uploads the PDF to AIPDF, then gets back a cloud link to paste into *any* ChatGPT chat. User opens chat, pastes cloud link once, link registers and user can begin to ask questions.

How to Use:
A question is asked, ChatGPT accesses document in cloud link, reviews document, pulls relevant information then generates a response

PDF

Project Manual

Ai PDF

Cloud Link

LLM Black Box

Question

Question

Question

Answer

LLM Black Box

All info related to question
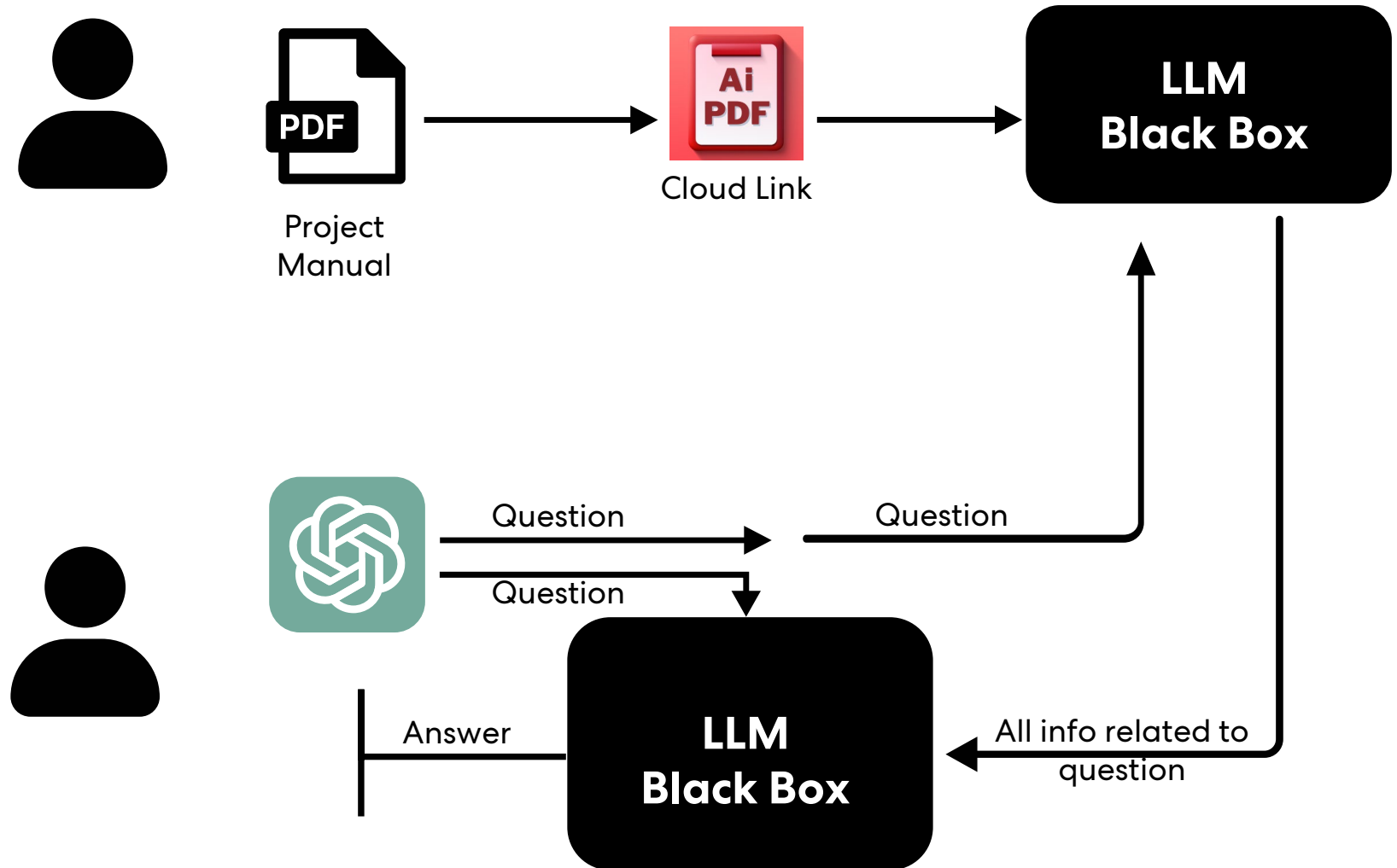
Research Questions

After project interviews while we were reviewing the various commercially available tools, we had three main questions that we wanted to answer as a part of this innovation incubator to really test AI enhanced projects specifications:

1. Will AI Enhanced project manual review help yield faster results when answering submittals?

2. Will AI Enhanced project manual review offer a better user interaction between designers/architects and project manuals?

3. Will AI enhanced project manual review assist younger designers who are in-experienced in traversing project manuals?

User Study Design:

In order to test questions related to our hypotheses, our initial step involved creating a simulated project manual. We collaborated with the specification writer we had interviewed for this purpose. Our aim was to include sections that are commonly found in project manuals across various projects. After compiling our mock project manual, which spanned 100 pages, we incorporated real submittals from an active healthcare project in the Boston Studio. This process was carried out in conjunction with the insights gathered during our interview with Perkins&Will's General Counsel. Confidential and sensitive information was meticulously removed, leaving predominantly product-related data that required review.

We tasked users with completing three submittals using a variety of tools. The first submittal served as a training exercise, helping users become comfortable with framing questions and prompts for ChatGPT enhanced with AI PDF capabilities. The second submittal involved partial product review, with users required to assess half of the products using ChatGPT and the remaining half using Bluebeam Review. The final submittal was designed as a submittal disguised as a substitution request, a Scenario that commonly occurs in the field more frequently.

Submittal #07 42 13 -4 – Aluminum Composite Wall Panels

Submittal #09 29 00-0.0 - Gyp Bd Auxiliary Materials

Submittal #03 30 00-S-2.8 A1 - Vapor Barrier

User Study Design:

We aimed to include approximately 30 users in our study but ultimately received participation from 20 individuals within Perkins&Will who were willing to engage. To encourage participation, we provided an incentive in the form of a $20.00 Amazon gift card. On the date and time of the User Study, users were sent, a Google Forms link for them to complete the study, a ChatGPT login and password, preloaded chat sessions with the mock specifications, and a prompt cheat sheet for those who might have been less familiar with AI tools, offering guidance on formulating effective questions and prompts.

Photo of the user study google forms

Photo of the user study google forms

Photo of the user study google forms

The tasks
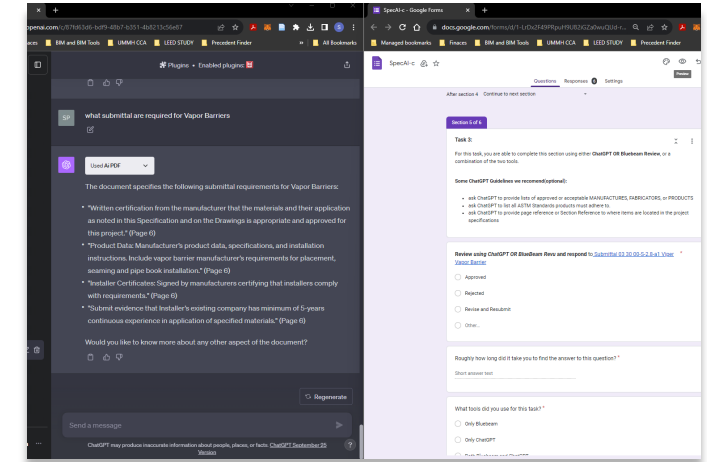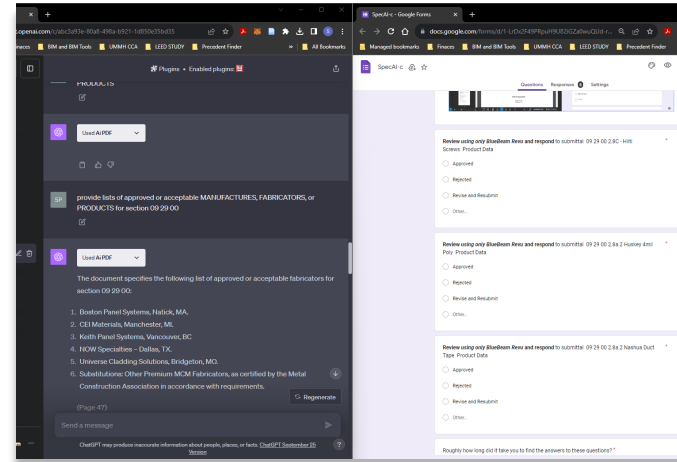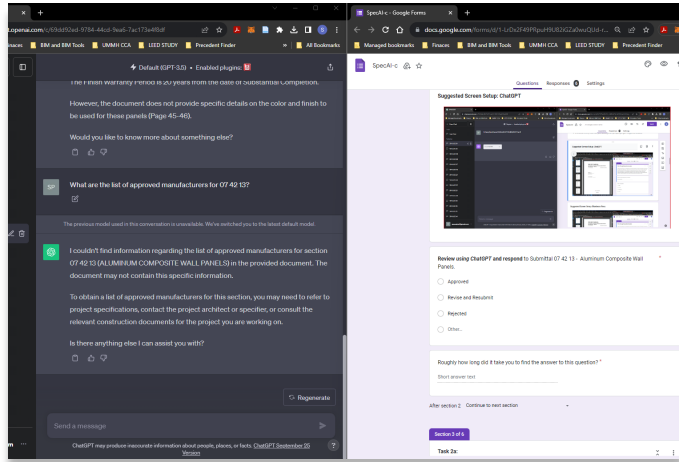
Training

The Training task was set up to allow users to become more comfortable with prompting ChatGPT about the project specifications. The training task asked users to review and approve an aluminum composite wall panel product. *This task was not recorded or included in our final study results.*

Task 1 (condition a, b):

For this task, we had a part submittal with 6 Auxiliary Gypsum board products. (3) were to be answered using ChatGPT, (3) were to be answered using traditional Bluebeam workflows. *We dictated in the submittal what tools needed to be used on which products and created 2 conditions.*

Task 2:

This last task was a substitution request disguised as a submittal for a vapor barrier product. This is something that happens more than we would like in project work. *We allowed users to use ChatGPT or Bluebeam Revu or combination of both tools to complete the task.*

User Study Design:

Users were given a recommended screen setup to optimize their efficiency during the study. However, it was observed that most participants chose not to utilize this format. Nevertheless, those who did follow the suggested screen setup appeared to experience significant time savings when reviewing both ChatGPT and Bluebeam.
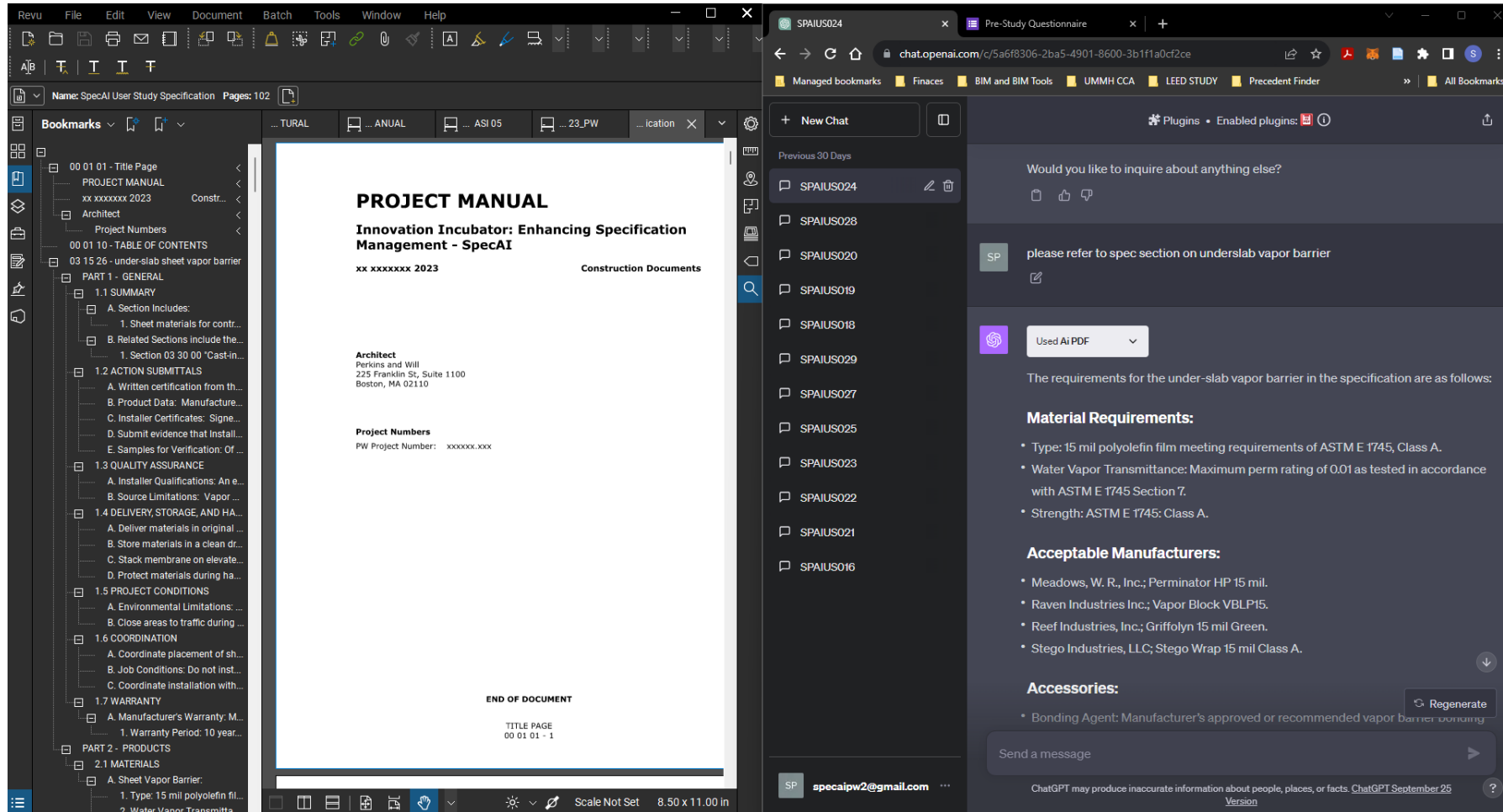
Photo of recommended screen setup for the study

Results

User Study:
User Demographics

Before we had folks jump directly into the user study, we had them fill out a pre-study Questionnaire to capture some demographic information. Some of that can be seen here:

Age of Participants
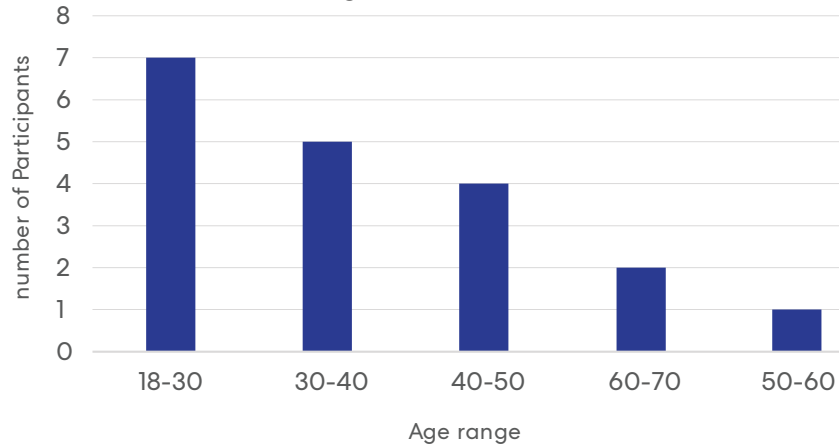
Genders of Participants:

■ Female   ■ Male   ■ Non Binary

Total Participants: 20
Female: 7
Male: 13
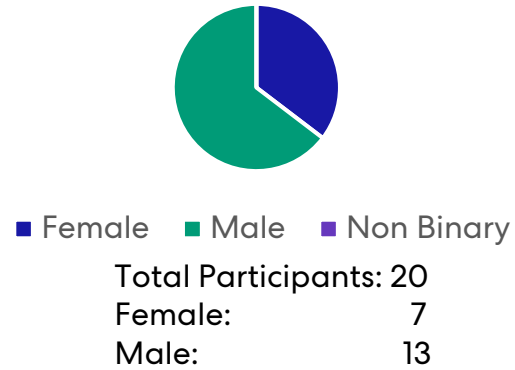
User Chat GPT Skill Level
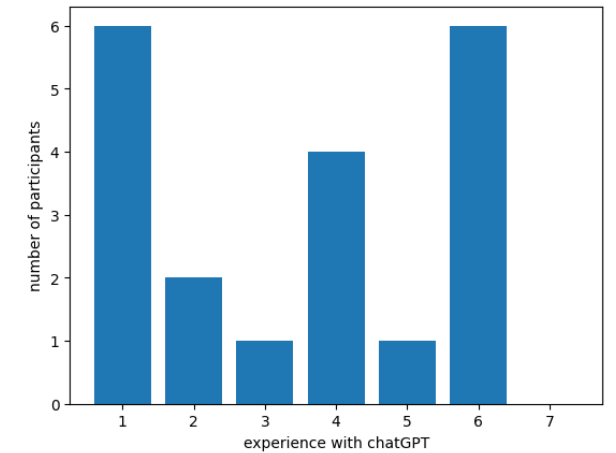
Roles of Participants
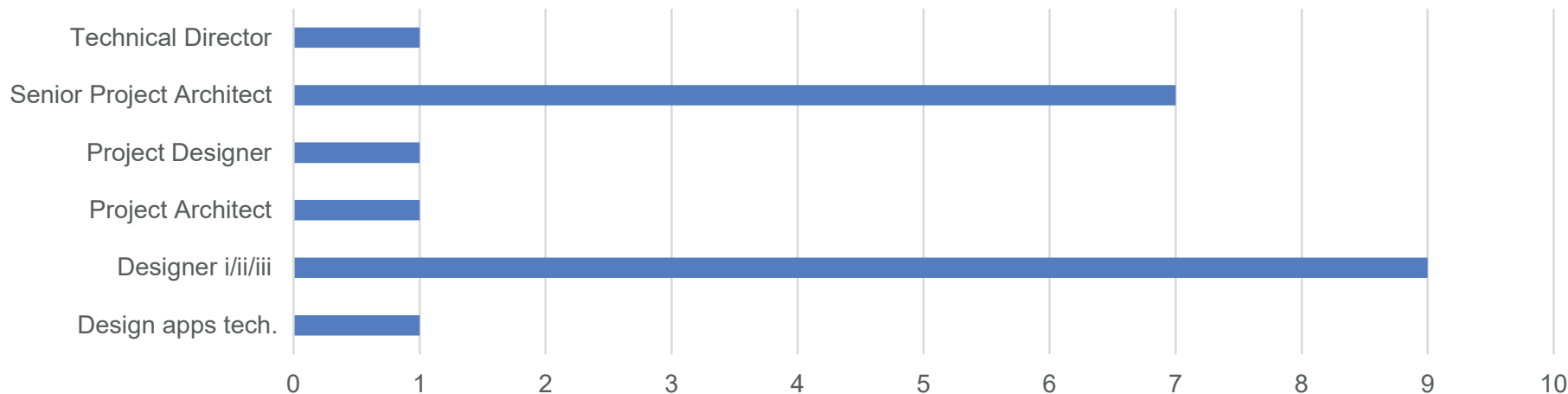
User Specification Skill level
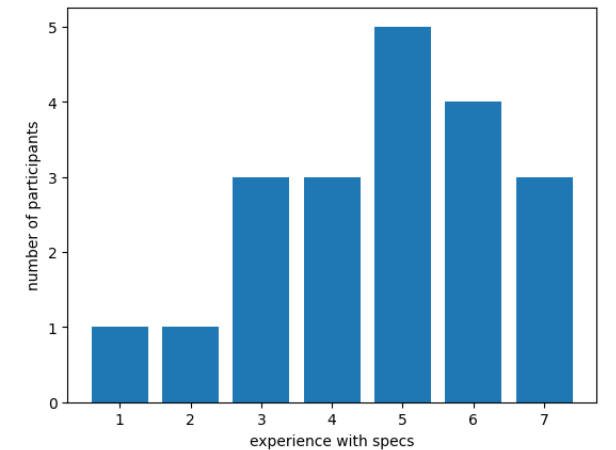
User Study:
Task 1

In the first task, participants responded to two groups of questions. Each group contained three submittals and was assigned one of the two tools: Bluebeam (control), or the customized ChatGPT workflow. The tool assignment and the order of the question groups was randomized.

We recorded the response to each submittal (Accept, Reject, Revise and Resubmit), and the self-reported time and confidence when using each of the two tools.

To analyze the results, we counted the number of correct answers given with each tool (minimum: 0, maximum: 3) shown in Figure 1. The medians of the two groups were the same, but results using ChatGPT were more spread overall (Figure 3).

We do not observe a significant difference in the time spent using each tool (Figure 3).
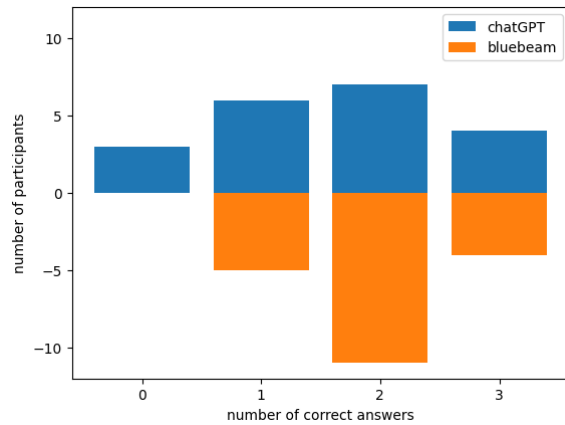
Figure 1
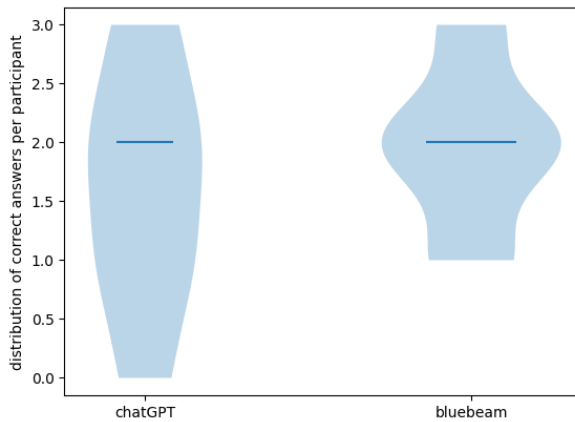
Figure 2
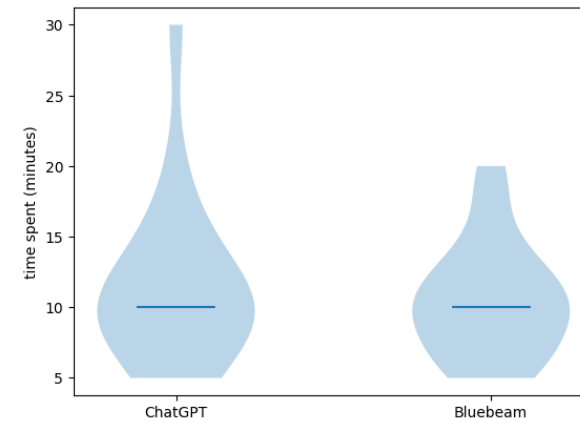
Figure 3

User Study:

Task 1

Participants felt more confident when using their standard tools instead of the suggested ChatGPT workflow.
Participants did not have a difficulty finding the relevant information for answering the submittal when using ChatGPT.

I feel confident about my answers

Finding the information was easy.
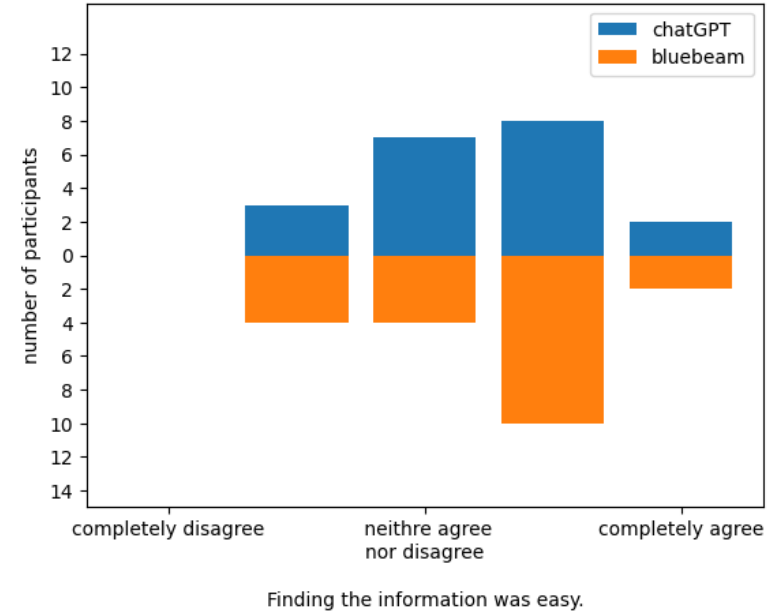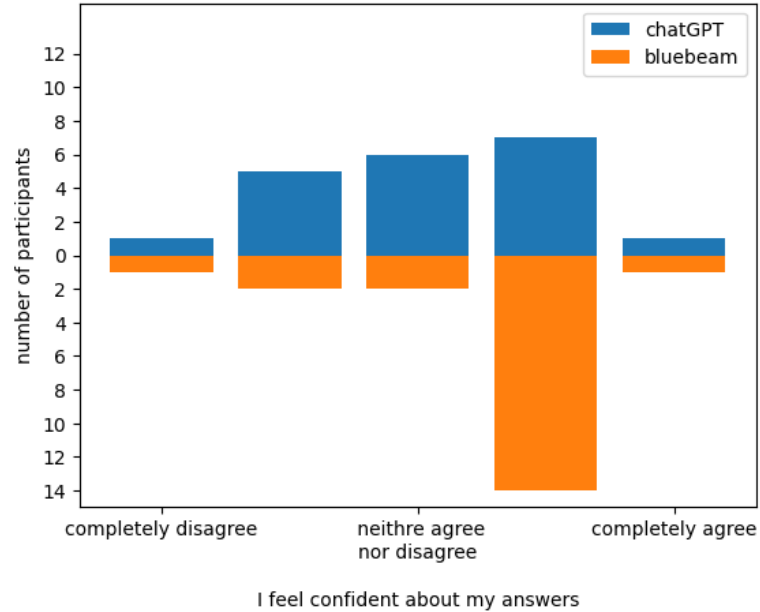
Correct answers by Spec experience

To better understand how the AI-based tool may have affected participants with different levels of prior experience with specifications, we further analyzed the results by splitting the participants in two groups based on self-reported experience (Figure 1).
First, we identified participants with little spec experience (experience < 4) and participants with high spec experience (experience >4). Figure 2 shows that when using ChatGPT, participants with higher experience achieved a higher median score. In addition, participants with little spec experience achieved a higher score when using Bluebeam.
Surprisingly, when the gap of experience between the two groups was increased (little experience < 3, high experience > 5), the inverse trend emerged (Figure 3). Participants with very little experience did best when using ChatGPT.
It must be noted, that due to the small sample and effect sizes none of these observations are statistically significant, and the interpretation of these results is mostly speculative.

Figure 2
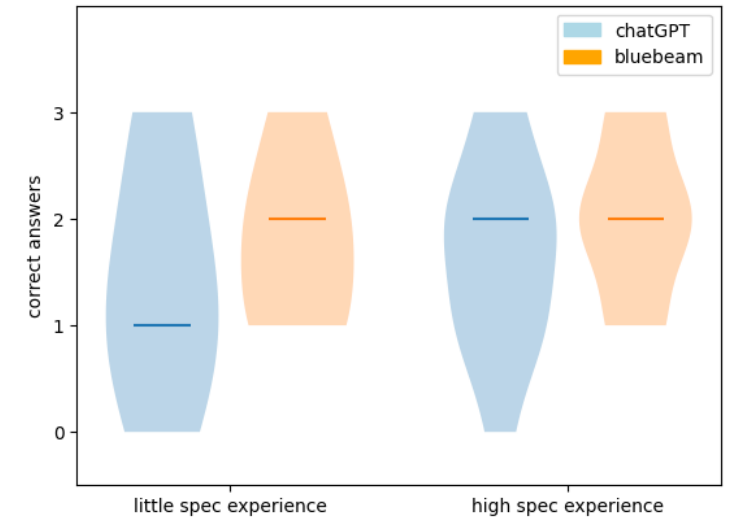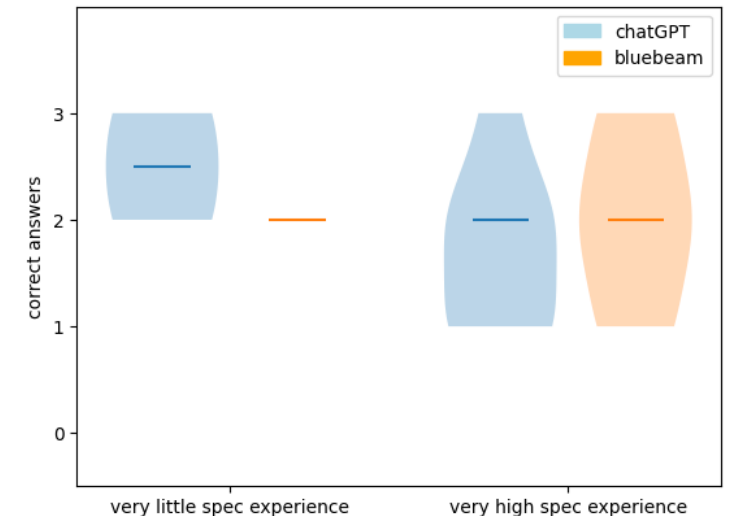
Figure 1

Figure 3

Correct answers by ChatGPT experience

A similar visualization of the data by ChatGPT experience does not indicate an effect of experience with ChatGPT prior to the study to the way that participants answered.

Figure 2
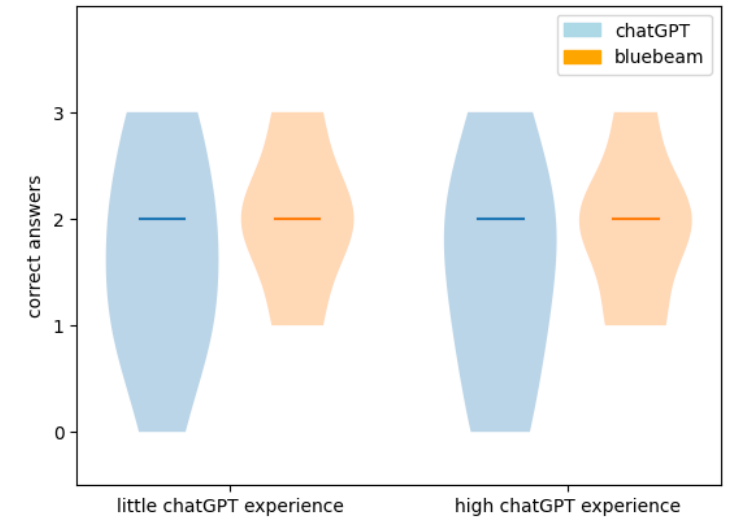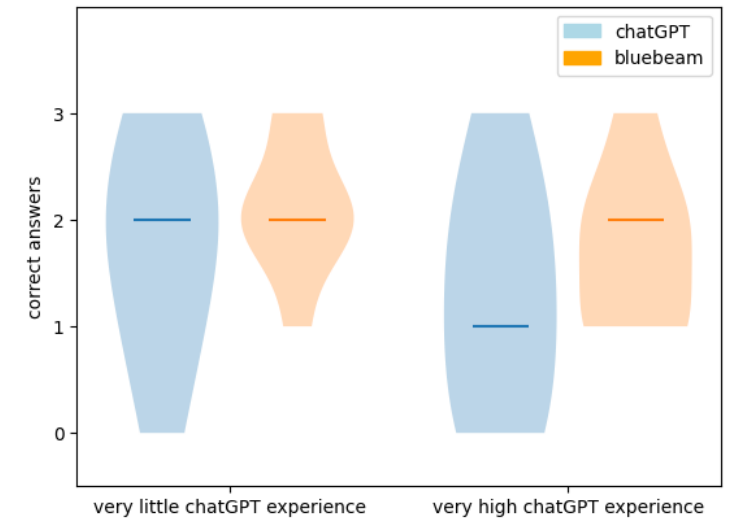
Figure 1

Figure 3

Time

Figures 1 and 2 illustrate the self-reported time spent using each of the two tools against the number of products that were correctly reviewed.
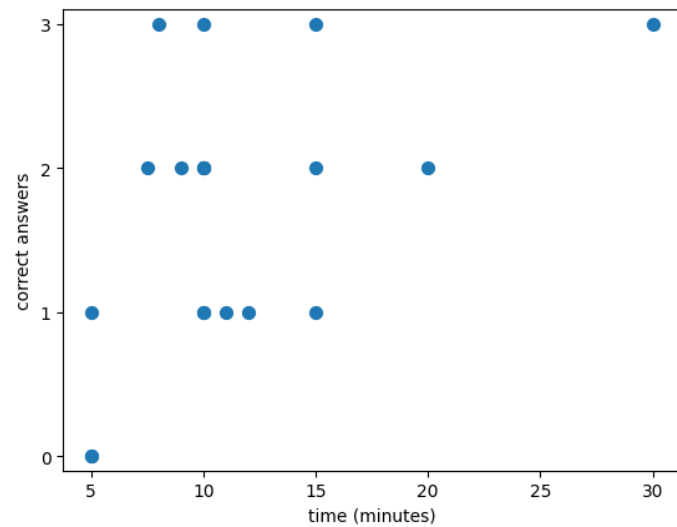
Figure 1: Condition **ChatGPT**
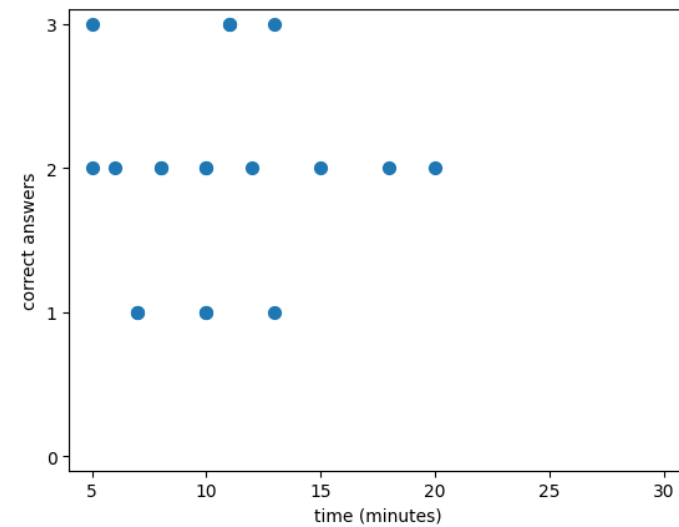
Figure 2: Condition **Bluebeam**

User Study:

Task 2

In the second task, the participants were asked to review a product submittal using any of the available tools (Bluebeam, ChatGPT). The majority of the participants chose to use both tools (Figure 1). A majority of participants responded that ChatGPT was the most helpful tool for answering this question. All three groups of participants - using Bluebeam, ChatGPT, and the combination- produced responses with the same average accuracy (Figure 2).
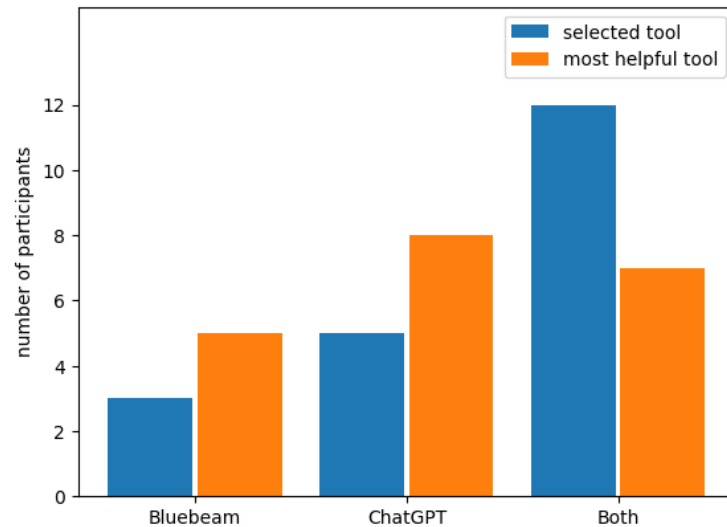
Figure 1
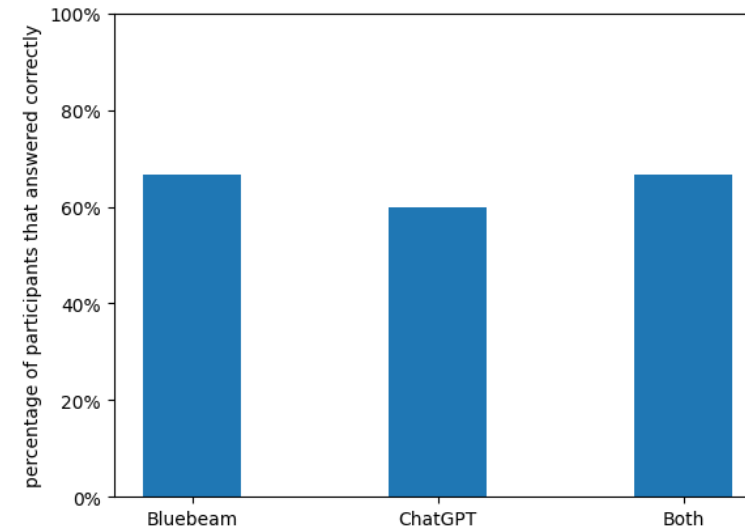
Figure 2

Post-study questionnaire

Most participants enjoyed using ChatGPT for submittal reviews. Although not many people were dissatisfied, the general satisfaction seems lacking, which is on par with the ease of finding information and the self-reported confidence.

Most participants agreed that the ChatGPT workflow is an improvement to current processes, and almost all participants responded that they would likely use, given the option.

ChatGPT offers an improvement compared to our standard tools.

How likely would you use the AI tool if it was approved by the company?
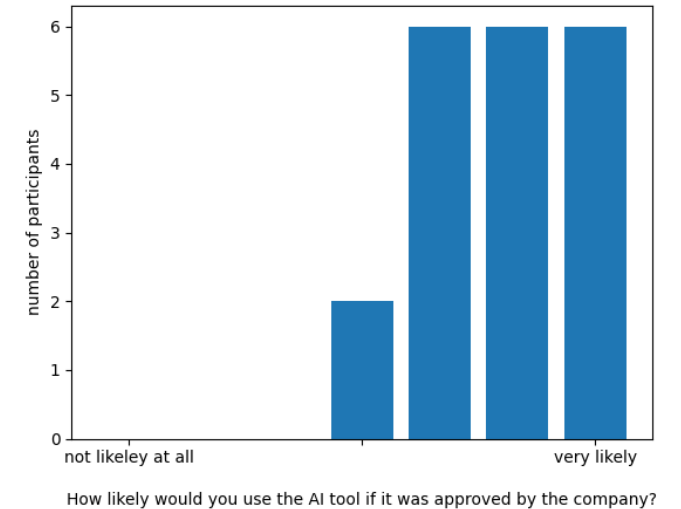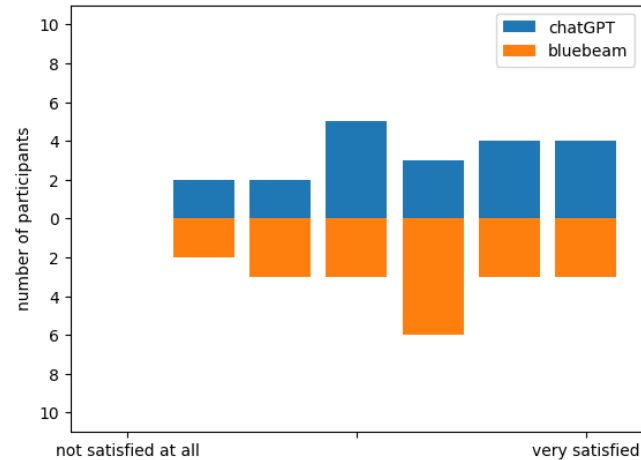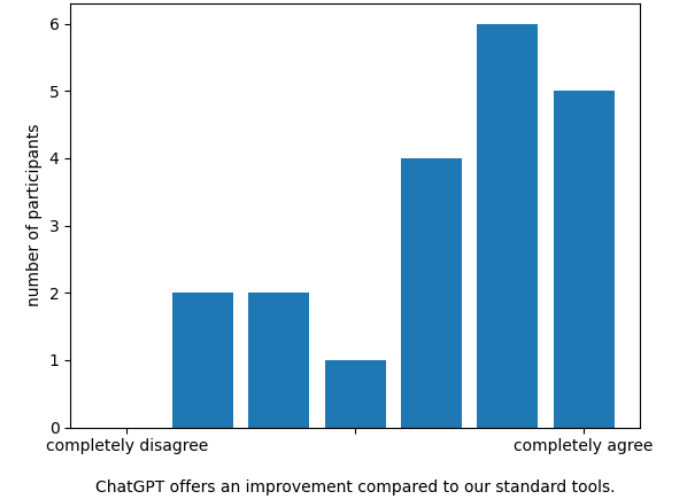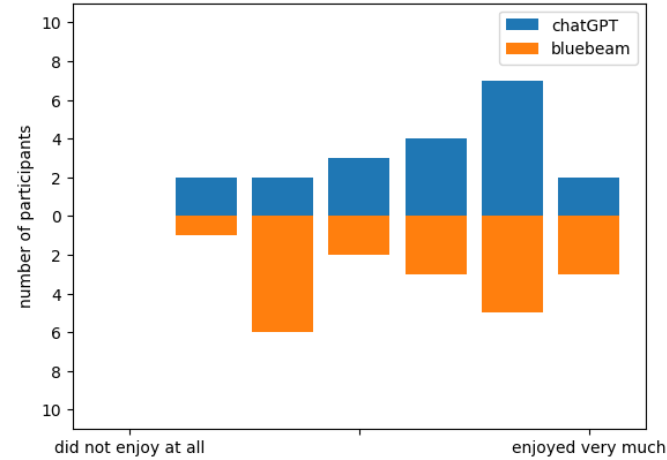
Research Questions: evaluation

After project interviews while we were reviewing the various commercially available tools, we had three main questions that we wanted to answer as a part of this innovation incubator to really test AI enhanced projects specifications:

1. Will AI Enhanced project manual review help yield faster results when answering submittals?

2. Will AI Enhanced project manual review offer a better user interaction between designers/architects and project manuals?

3. Will AI enhanced project manual review assist younger designers who are in-experienced in traversing project manuals?

Discussion

- Participants eager to augment spec review with AI.

- Need for structured tool (people did not follow most convenient setup, chatgpt had to be conditioned, prep for users should be in place)

- Time was the same but people may get faster with ChatGPT (new tools, learning curve)

- Provide submittal to AI (by participants' request)

Discussions:
Lessons Learned

- Oversubscribe

- Anticipate technical failures (timer, account lock, …)

- Constrain Interface

Conclusions:

- ChatGPT cannot used out-of-the box for such specialized task.

- Conversational AI **can** improve user interaction around submittals.

- Accuracy of the tool needs to be improved.

- Ability to handle and correctly reference multiple documents could be useful (access to submittal).

- Custom-built tool for the task vs compilation of general-purpose products is what will be needed.

Bibliography

Research:

Bagchi, Mayukh. **"Conceptualising a Library Chatbot Using Open Source Conversational Artificial Intelligence."** *DESIDOC Journal of Library & Information Technology*, vol. 40, no. 06, Dec. 2020, pp. 329–33, https://doi.org/10.14429/djlit.40.06.15611.

Bello, Abdul-Lateef, et al. **"*Framework Development for the Selection of Cloud Services for Construction Industry – A Case of Conversational-BIM Chatbot.*"** 4308361, 28 Dec. 2022, https://doi.org/10.2139/ssrn.4308361.

Bielski, Jessica. *The Morphological Echo of Architects. Concept for a Conversational Artificial Intelligence to Support Architects during the Early Design Stages*. 2022.

Dwivedi, Yogesh K., et al. "'**So What If ChatGPT Wrote It?' Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy."** *International Journal of Information Management*, vol. 71, Aug. 2023, p. 102642, https://doi.org/10.1016/j.ijinfomgt.2023.102642.

Gkinko, Lorentsa, and Amany Elbanna. **"The Appropriation of Conversational AI in the Workplace: A Taxonomy of AI Chatbot Users."** *International Journal of Information Management*, vol. 69, Apr. 2023, p. 102568, https://doi.org/10.1016/j.ijinfomgt.2022.102568.

Kulkarni, Pradnya, et al. **"Conversational AI: An Overview of Methodologies, Applications & Future Scope."** *2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, 2019, pp. 1–7, https://doi.org/10.1109/ICCUBEA47591.2019.9129347.

Saka, Abdullahi B., et al. **"Conversational Artificial Intelligence in the AEC Industry: A Review of Present Status, Challenges and Opportunities.**" *Advanced Engineering Informatics*, vol. 55, Jan. 2023, p. 101869, https://doi.org/10.1016/j.aei.2022.101869.

You, Hengxu, et al. **"*Robot-Enabled Construction Assembly with Automated Sequence Planning Based on ChatGPT: RoboGPT.*"** arXiv:2304.11018, arXiv, 21 Apr. 2023, https://doi.org/10.48550/arXiv.2304.11018.

Graphics and images:

https://askyourpdf.com/

https://www.chatpdf.com/

https://pdf.ai/

https://filegpt.app/

https://aipdf.app/

https://bard.google.com/

https://chat.openai.com/

https://www.bing.com/