# Human-AI Co-Creation of Worked Examples for Programming Classes

Mohammad Hassany<sup>1</sup>, Peter Brusilovsky<sup>1</sup>, Jiaze Ke<sup>2</sup>, Kamil Akhuseyinoglu<sup>1</sup> and Arun Balajiee Lekshmi Narayanan<sup>1</sup>

<sup>1</sup>University of Pittsburgh <sup>2</sup>Carnegie Mellon University



#### **MOTIVATION**

- Traditional form of code-explanations examples focus on passive learning
- Motivated researchers create interactive tools
- Although instructors can reuse other examples, they prefer to use their own examples
- Creation of an explained example takes a lot of time

## RELATED WORKS WORKED EXAMPLES

- In WebEx [1], students explore line-by-line instructor-provided comments
- PCEX [2] elevates example study process to the ICAP's interactive and constructive levels
- PCEX is currently the most detailed approach for explaining worked examples

#### LLMS FOR CODE EXPLANATION

- GPT-3, 3.5, 4, OpenAI Codex, Github Copilot are used to generate code-explanations at different levels: line-by-line, step-by-step, and high-level summary
- LLM-generated explanations are evaluated by authors, students, and tool-users
  - Useful, easier, and more accurate than learner-sourced explanations
  - o Mistakes can be corrected by authors
- This is an opportunity for Human-AI Collaboration

#### THE FEASIBILITY STUDIES

- 1. A well-formed prompt based on research and internal evaluations findings
- 2. ChatGPT and Experts explanations are twice the size of students explanations
  - Despite the length difference, experts and students' explanations have very similar readability; also, more similar with each other than ChatGPT explanations
  - ChatGPT explanations are much less readable
- ChatGPT-generated explanations are rated as more complete and judged to be better than experts explanations (but not with a clear win)
- ChatGPT-generated explanations benefit from experts review before presentation to students

### **WORKED EXAMPLE AUTHORING TOOL - WEAT**

- Authors can utilize ChatGPT for explanation generation (optional)
- Generated explanations can be reviewed, refined, and removed if necessary
- Authors can also provide more explanations
- An example of Human-AI co-creation: ChatGPT handles time-consuming task, while human expert has the control to accept, refine, or reject

#### **EVALUATION**

- Engaged 5 instructors (A1-A5) who teaches Java and Python introductory classes
- Video tutorial and user manual were prepared and shared with instructors
- Instructors' interactions were logged and used for analysis
- Half of generated explanations used with no change
- An overall Levenshtein-ratio of 0.73 for created worked examples' explanations

	A1	A2	A3	A4	A5	Total
Examples Created	6	2	2	1	1	12
Generated Explanations	126	32	44	8	27	237
Lines of Code being Explained by ChatGPT	55	12	21	2	9	99
Explanations Excluded	18	2	4	0	0	24
Explanations Liked	6	0	9	0	0	15
Explanations Edited	29	0	11	0	26	66
Explanations Removed	8	0	15	0	0	23

#### **CONCLUSION**

- Although ChatGPT-generated explanations were rated positive, in some case, they were not as good as experts' explanations
- Experts explanations are closer to students active vocabulary - ChatGPT explanations benefit from experts review before presentation to students
- A considerable amount of worked example's explanations are ChatGPT generated - saving time and effort

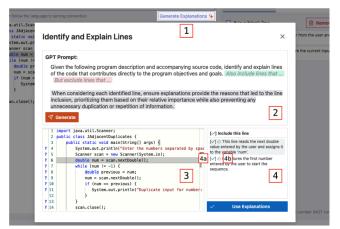


Figure 3: Human-Al Collaborative Worked Example Authoring, 1) "Generate Explanations" button, 2) default prompt (author can tune the prompt - optional), 3) program source preview, 4) generated explanations for the selected line.

#### **REFERENCES**

- Brusilovsky, P., Yudelson, M., & Hsiao, I.-H. (2009). Problem Solving Examples as First Class Objects in Educational Digital Libraries: Three Obstacles to Overcome. Journal of Educational Multimedia and Hypermedia, 18(3), 267–288.
- Hosseini, R., Akhuseyinoglu, K., Brusilovsky, P., Malmi, L., Pollari-Malmi, K., Schunn, C., & Sirkiä, T. (2020). Improving Engagement in Program Construction Examples for Learning Python Programming. International Journal of Artificial Intelligence in Education, 30(2), 299–336. https://doi.org/10.1007/s40593-020-00197-0



