Leveraging Human-AI Collaboration in Crowd-Powered Source Search: A Preliminary Study

Yong Zhao, Zhengqiu Zhu, Bin Chen, and Sihang Qiu*

Abstract: Source search is an important problem in our society, relating to finding fire sources, gas sources, or signal sources. Particularly, in an unexplored and potentially dangerous environment, an autonomous source search algorithm that employs robotic searchers is usually applied to address the problem. Such environments could be completely unknown and highly complex. Therefore, novel search algorithms have been designed, combining heuristic methods and intelligent optimization, to tackle search problems in large and complex search spaces. However, these intelligent search algorithms were not designed to address completeness and optimality, and therefore commonly suffer from the problems such as local optimums or endless loops. Recent studies have used crowd-powered systems to address the complex problems that machines cannot solve on their own. While leveraging human intelligence in an AI system has been shown to be effective in making the system more reliable, whether using the power of the crowd can improve autonomous source search algorithms remains unanswered. To this end, we propose a crowd-powered source search approach enabling human-AI collaboration, which uses human intelligence as external supports to improve existing search algorithms, and meanwhile reduces human efforts using AI predictions. Furthermore, we designed a crowd-powered prototype system, and carried out an experiment with both experts and non-experts, to complete 200 source search scenarios (704 crowdsourcing tasks). Quantitative and qualitative analysis showed that the sourcing search algorithm enhanced by crowd could achieve both high effectiveness and efficiency. Our work provides valuable insights in human-AI collaborative system design.

Key words: source search; crowdsourcing; crowd-powered systems; crowd computing

1 Introduction

Source search problems always exist in nature and in our daily lives, such as animals finding an odor source to acquire foods in the wild and people searching for the emission source of air pollution. Traditional search algorithms, such as tree search algorithms and graph search algorithms, work well for limited search space

given enough time. However, as technology advances and computing power explodes, people started to expect search algorithms to solve search problems in more complex scenes with a more strict time limit.

Complex search space could be high-dimensional and dynamic, and the computation might be required to be completed in real-time in many applications. Since traditional traverser algorithms can no longer meet those requirements, novel search algorithms have been proposed to tackle source searching problems in large and complex search space [1, 2]. These novel search algorithms integrate human cognition and animal behaviors into the search rules, and gather information during the search process to dynamically adjust search parameters. While search efficiency has been dramatically im-

● Yong Zhao, Zhengqiu Zhu, Bin Chen, and Sihang Qiu are with the College of Systems Engineering, National University of Defense Technology, Changsha, China. E-mail: {zhaoyong15, zhuzhengqiu12, chenbin06, qiusihang11}@nudt.edu.cn

* Corresponding author.
  Manuscript received: year-month-day; accepted: year-month-day

proved, completeness and optimality can no longer be guaranteed. Therefore, it is possible that these search algorithms return a local optimum or even no solution, instead of the global optimal solution. Researchers and practitioners have noticed this issue [3].

Recent work has focused on crowd-powered systems and human-AI collaborations [4]. These approaches enable humans to take part in the automatic process that is supposed to be completely controlled by AI or automatic machines to solve complex problems. A crowd-powered system provides us with a new perspective that humans could get involved in the automatic problem-solving process, to enhance the effectiveness and efficiency of algorithms [5–8]. Since human-AI collaboration has been proven to be feasible in a variety of domains[9], we see an opportunity that combines human rationales with search algorithms, to overcome the difficulties that current source search approaches usually encounter. However, whether the power of the crowd is really effective and efficient in improving existing source search algorithms remains unknown. To address this knowledge gap, in this work, we are particularly interested in answering the research question: *how can crowd-powered approaches improve the effectiveness and efficiency of source searching algorithms?*

To answer the research question, we designed a human-AI collaborative framework that could improve the existing source searching algorithm in a way that crowdsources the problems that occurred during source searching. We implemented a prototype system where a virtual robot autonomously searches a source in complex environments in a simulation setup, to enable a user study. The crowd-powered system is responsible for detecting fatal problems, explaining the algorithm, giving suggestions, and generating tasks for humans to complete. When it comes to humans' turns, humans could either take full control of the robot or aid the robot to address problems. Particularly, to better facilitate effective problem-solving, the system predicts the location of the source using Bayesian methods and sequential Monte Carlo methods to further assist humans in making decisions and taking actions. We recruited 10 participants in this study, including 4 domain experts in the field of source searching, and 6 non-experts having no experience in source searching, to evaluate the proposed human-AI collaborative approach in randomly generated complex searching environments.

In total, 200 source search scenarios (704 crowdsourcing tasks) were completed by the collaboration of humans and autonomous source search algorithms. The experiment shows that the crowd-powered system could improve the performance of the state-of-the-art source searching algorithms, in both effectiveness (success rate 100%, 22% higher than non-human methods) and efficiency (using significantly fewer iterations/steps to find the source). Furthermore, we analyzed system usability scores and cognitive workload scores reported by the participant. Results show that a specific way of interaction could achieve better usability and cognitive workload for a group of participants in the prototype system. Our work provides useful suggestions, valuable insights, and important implications in leveraging human-AI collaboration for improving source search algorithms.

An earlier version of this work was published at ChineseCSCW 2022 and received the Best Student Paper Award. In this version, three main additional extensions were developed: 1) we further categorized the problems that autonomous algorithms could encounter during source search; 2) we further presented the necessary elements for effective human-AI collaboration; 3) we extended and formatted the post-task survey and analyzed participants' feedback systematically.

2 Related Work

We discuss related literature from two perspectives: crowd-powered systems and source searching.

2.1 Human-AI Collaborative Systems

The goal of designing a human-AI collaborative system is to leverage human rationales in a way combing with computer systems to collaboratively solve complex problems. Scientists at Microsoft proposed human-AI interaction guidelines, to help researchers and practitioners working on the design of studies and applications that use AI technologies [4]. The guidelines give 18 specific items categorized into four main parts, namely initially, during interaction, when wrong, and over time, to help people working in the AI and Human-Computer Interaction (HCI) communities better design and evaluate human-AI collaborative systems. Similar works have been done to provide important advice for better user experience in explainable AI and human-AI decision-making [10, 11]. A typical collaborative system is the ESP game [12], which is an image labeling system developed by Google. This system successfully gamified the image labeling process, and produced a large amount of data while the users were enjoying the

game. CrowdDB is another example leveraging human input to process queries that database systems cannot answer [5]. Bozzon et al. proposed Crowdsearcher, a system that is able to answer search queries using the intelligence of crowds [6]. Furthermore, previous work combined human intelligence with machine learning methods to address the problem such as conversational agent learning intents and text classification [9, 13]. Recent studies recruited online users from crowdsourcing platforms, and applied smart task scheduling and output prediction methods to produce city maps [14, 15]. While human-in-the-loop systems have been shown to be effective in many domains, algorithm-in-the-loop systems also started to play a critical role in human decision-making. Previous work introduced this concept, and provided principles for human-AI decision-making and risk analysis [16, 17]. In the domain of robotics and engineering, human-AI collaboration has been used for a long time, to address practical problems that can hardly be considered in theoretical models [18, 19]. For instance, human-AI collaboration was effectively applied to address radiation source search and localization [20], spill finding and perimeter formation [21], and urban search and rescue response [22].

2.2 Source Search

In general, source search is a kind of problem that aims to determine the location of the source (of gas or signal) in the shortest possible time, as it is of vital importance for both nature and mankind [23–25]. For example, the search for preys [26], submarines [27] survivors [28], and pollution sources [29]. As a classical kind of source search algorithm, the bio-inspired algorithm typically leverages the gradient ascent strategy to approach the source, based on a reasonable assumption that the signal emitted by the source has a greater intensity near the source [30, 31]. However, in the presence of environmental disturbances (e.g., turbulence), the intensity gradient of the emitted signal may be disrupted, undermining the feasibility of the bio-inspired searching algorithm [2]. An alternative kind of source search algorithm has been developed based on Bayesian theory [23]. Previous works [1, 2] proposed the cognitive search algorithm that models the source search process as a Markov Decision Process. To further enhance the performance (i.e., success rate and efficiency) of a search algorithm, multi-robot collaboration mechanisms [32–34] were designed and adopted. However, when source search happens in complex environments,

the search process always encounters fatal problems, resulting in wrong outcomes. In this work, we designed a prototype system for source search in complex environments, and carried out a user study with this system to answer the research questions.

3 Methods

In this section, we propose a method that leverages human-AI collaboration to improve existing autonomous source search algorithms. We organized a discussion with three experts (including two authors of this work) in the domain of source search, with the aim of answering the following three questions:

(1) What fatal problems could be happened during the search process?

(2) What kinds of information are necessary for humans to understand the problem that the search process is having?

(3) If a search algorithm is troubled by a problem mentioned before, how could the problem be addressed by humans?

According to the discussion, we designed a crowd-powered framework that combines human intelligence with AI, to overcome the problems that current source search algorithms often encounter. We classified the common problems that could occur during the search process, and showed how these problems could be understood and addressed by crowds from experts' perspectives. We further explain each step and elaborate on how humans and AI play their own roles within the proposed framework.

3.1 An Overview

In this section, we show the design of a crowd-powered source search method, and explain how human rationales can be used during the search process to improve the effectiveness and efficiency of search algorithms. There are numerous ways to enable crowd-powered methods in search, and human-AI hybrid intelligence can naturally play a role in each part of a search algorithm. In this study, we designed this simple framework in a way to minimize human efforts, and try not to make changes to the mechanism of source search algorithms. The overview of the method is shown in Fig. 1.

The entire workflow consists of three main steps. The first step – initialization, defines the search goal, space,
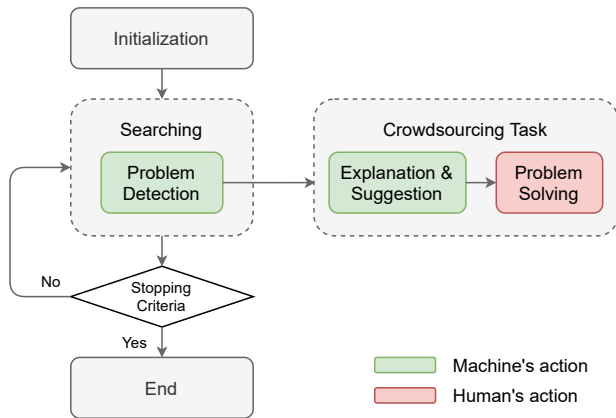
Fig. 1 The crowd-powered method that integrates human-AI collaboration into the search process.

rules, and parameters. In the second step, the framework conducts the source search algorithm using the rules and parameters defined in the first step, to search for the source in the search space. When the search stopping criteria are satisfied (e.g., the source is found or the space has been fully searched), the workflow ends and outputs the corresponding result.

3.2 Detecting Problems

In the proposed framework, to minimize human effort, the algorithm is taking care of the search process most of the time, while human interaction is enabled when an algorithm is having a problem that cannot be effectively addressed on its own. Therefore, in order to detect such problems and enable human-AI collaborations, during the search process, the AI system needs to automatically monitor and acquire the state of the search process.

According to the discussion with experts to answer the question of "*what fatal problems could be happened during the search process*", we classified common problems found in search algorithms, especially in informed search algorithms or other intelligent search strategies which cannot guarantee the completeness and optimality [3].

(1) **Local optimum**: this problem could happen when the search space is very large or infinite, where traditional search algorithms cannot efficiently work. In this case, a search algorithm might find a local optimum and believe this is the final goal.

(2) **No information gained**: this problem means that there is no new information gained while the search state is kept updated. It could happen to informed search or informative path planning strategies that

always require new information to guide the search direction. A wrong search direction can be fatal when the search space is very large or infinite.

(3) **Dead end**: this problem could happen when searching heuristics, cost/reward functions, or stopping criteria are not appropriately set up for search environments. In this case, the search algorithm might end up with a dead-end, and it returns no solution in the end.

(4) **Endless loop**: this problem could happen to local or random search strategies, which do not record all the previous search states. If the search heuristics or cost/reward functions are not correctly defined, the algorithm may get trapped in endless loops.

Problem detection could be achieved in various ways. A simple solution would be using a rule-based expert system to automatically detect the pattern of search history. It is also possible to train a machine learning model to detect the pattern of problematic search states. In this preliminary study, we simplified the problem detection part, which does not use machine learning to recognize problem patterns. According to experts' domain knowledge, for current source searching algorithms, local optimum or dead-end usually strongly relates to no information gained and endless loop problems eventually. Therefore, we proposed a simple rule-based mechanism to detect the no information gained and endless loop problems automatically: if the searcher 1) passes by the same spot 5 times within a specific time window, and 2) acquires zero measurement, the system stops the search process.

3.3 Crowdsourcing Tasks

When a problem is detected, the method automatically generates crowdsourcing tasks to collect and leverage human intelligence. The key steps for successful task completion are considered to be task explanation and problem-solving. The AI behind the autonomous source search algorithm is responsible for task explanation and for giving valuable suggestions, while crowd workers (humans) are responsible to figure out the problems that the algorithm cannot solve.

3.3.1 Task Explanation Using Artificial Intelligence

The explanation is a very important part of human-AI collaboration. When the algorithm is having a prob-

lem, humans need to first understand how the algorithm works and what the problem is, then they can help the AI address the problem. In this study, explaining source search algorithms and problems is slightly different from explaining machine learning models, since machine learning models are usually black boxes and it is not easy for humans, even experts, to understand why a model makes a problematic prediction. In the context of source search algorithms, although how the algorithm works is clear to designers and experts, the actual search process (search states) usually remains hidden from humans. Therefore, AI systems could be used to give useful explanations and suggestions so that humans can easily understand what happens and get some tips to solve the problem.

With the question of "*what kinds of information do you need in order to know the problem that the search process is having*", we discussed with experts and explicitly asked for requirements of good explanations and suggestions. As a result, we presented the necessary elements in Fig. 2:
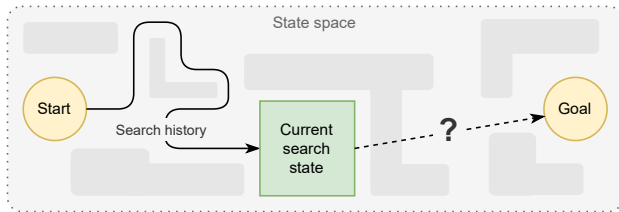
Fig. 2 Elements needed to be presented while explaining the source search algorithm.

(1) The AI system should present the current search state.

(2) The AI system should present the partial/entire search space.

(3) The AI system should present the partial/full history of the search process.

(4) The AI system should describe the problem of the current search state, and better show why it happens.

(5) The AI system could present the initial state.

(6) The AI system could suggest a solution.

The first three requirements are relatively easy to achieve, while the last three requirements need AI's actions. The AI system implemented behind the source

search algorithm should find a way to explain the problem in an understandable way, and further, the AI system could give an estimation of the source in the unknown environment, which would help the human better understand the problem of the current state and solve the problem later.

A variety of AI techniques could be used here to facilitate explanations and suggestions. In this study, the AI suggestion features a source estimation method that uses Bayesian inference and sequential Monte Carlo methods to show the distribution of the posterior probability of the source location (see green particles in Fig. 3) [35, 36]. The AI also suggests an area where the source most likely will be called "belief source area" using DBSCAN [37] to provide more information to help humans understand and address the problem.

3.3.2 Problem Solving Using Human Intelligence

After AI detects the problem and provides useful explanations and suggestions, it comes to the human's turn to address the problem that a search algorithm cannot handle on its own. With the question of "*If the search algorithm is really troubled by the problems you mentioned before, how would you address them*", we discussed with experts and summarized four main ways to overcome the problems, so as to improve the search algorithm:

(1) **Taking over the search process**. This way requires humans to take over the search process until the problem is addressed. It means that the search process is no longer running automatically, instead, humans decide what are the next search states to make the search proceed. This way requires a good understanding of the search algorithm and can provide maximum flexibility for humans to address the problem.

(2) **Manual pruning**. This way requires humans to prune the problematic searching branches for jumping out of the current problem, since some search paths could look obviously problematic to humans according to humans' experience and rationales, but these problematic search paths might seem correct to the AI.

(3) **Setting a temporary search goal**. This way requires humans to set up a short-term goal for the search process, to temporarily replace the end goal. To this end, the temporary goal lets the search process make a detour to bypass the problem. The

temporary search goal should be relatively easy to achieve, and should be "closer" to the end goal according to human understanding and reasoning.

(4) **Adjusting the search algorithm on the fly**. This way requires humans to adjust the parameters, functions, rules, or heuristics of the search algorithm on the fly to solve the problem. It also requires much experience in using such search algorithms, i.e., humans need to understand the setting, mechanism, and workflow of search algorithms very well, to address the problem effectively.

The key to using human rationales to improve source search algorithms is to find an appropriate way for the specific context, and design effective human-computer interaction means.

3.4 Prototype System Design

We designed a prototype system following the framework shown in Fig. 1.

3.4.1 Crowd-powered Infotaxis Algorithm

The source search algorithm used in this system is Infotaxis, one of the most popular novel search strategies particularly effective for source searching problems [1, 38].

When a problem is detected, a GUI-based crowdsourcing task is automatically generated. We use graphical elements to explain the algorithm as well as the problem, suggested by Fig. 2. In the prototype system, the goal of explanation is to let users clearly "see" the problem, rather than deeply "understand" the problem. We showed the direction that the robot (searcher) wanted to go, and the direction robot had to go (because of obstacles) to help people understand why a problem could happen. We did not further explain the reasons since a problem could be the consequence of many different factors. Future work could focus on a deep understanding of the problems.

When the human (a crowd worker) starts to operate, the crowdsourcing task provides two control modes. A full control mode allows the user to take over the search process and control every single step of the robot; An aided control mode allows the user to define a temporary goal (a targeted location), so the robot will pause the current search activities and move to the targeted location set by the user. We did not implement other problem-solving means in the prototype system since they require more expertise and incorrect operations may lead to a high failure rate. Future work could con-

sider implementing more control modes such as setting forbidden areas and search parameter tuning, representing manual pruning, and adjusting the search algorithm on-the-fly respectively.

3.4.2 Task Interface

As shown in Fig. 3, the task interface uses graphical elements to explain the search algorithm and the problem. Since the environment and the source are unknown at the beginning, on the interface, most areas except the cells near the robot are in black, and green particles (representing posterior distributions of the source location) are randomly distributed, as shown in Fig. 3 (a). After clicking the button [START], the search begins. When a problem is found, the system generates a task and gives the user a voice-based alert, and then the crowd worker can click on the interface with the button [EXECUTE] to execute the task, by controlling the robot or planning a path for the robot, as shown in Fig. 3 (b). The crowd worker can click the button [CONTINUE] to complete the task, letting the search process continues. When the source is found, the system shows a message saying "the source has been found successfully!" in red, then the user can click the [NEXT] button, to switch to the next source search scenario. The prototype system was developed using Python 3.7 and tkinter packages. The code repository of the prototype system is shared online publicly together with the data, for the benefit of the community[*].
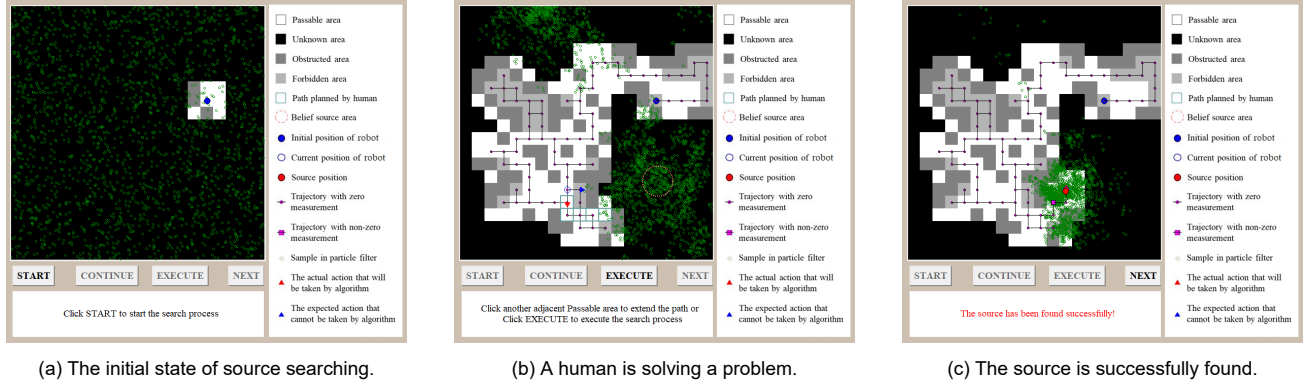
4 User Study

We design a user study to answer our research question. In this section, we introduce experimental conditions, environments, measures, and the procedure of the user study.

4.1 Experimental Conditions

As we have introduced in the previous section, we provided two interaction/control modes – Full Control (FC) and Aided Control (AC) respectively. A full control interaction mode represents the problem-solving method that requires humans to take over the search process, while an aided control interaction mode represents the problem-solving method that sets a temporary goal (let the robot exits the current search state and then navigate it to a manual defined location).

Furthermore, we used two baseline conditions in our experiment. The baseline 1 condition directly uses

---

[*]`https://github.com/PSimLab-Student/`
`human-AI-cooperation.git`

(a) The initial state of source searching.   (b) A human is solving a problem.   (c) The source is successfully found.

Fig. 3   Screenshots of crowdsourcing tasks generated by the prototype system.

the state-of-the-art source search algorithm (Infotaxis), while the baseline 2 condition also uses our proposed automatic problem detection method and then navigates the robot to a random location in order to jump out of the problem. Please note that the baseline 2 condition is also an improvement based on the state-of-the-art source search algorithm.

4.2   Experimental Environments

The source search activities are performed by a virtual robot in a 2D $20m \times 20m$ squared area in a simulation setup. The 2D search area is divided into 20×20 cells in a grid. Each cell has a probability P_o determining whether this cell contains an obstacle. P_o is set to be 0.75, to give a relatively high difficulty (more obstacles) of tasks, since simple environments (with few obstacles) do not need human assistance that much. In this study, we did not consider the specific types or shapes of obstacles. If there is an obstacle in a cell, the cell is considered to be completely obstructed and cannot be arrived at or passed by the robot.

The prototype system was deployed on a PC, and all the participants were invited to execute tasks using the same PC to ensure a fair comparison. Participants were invited to a quiet lab, to make sure the experiment would not be interrupted by others.

4.3   Measures

In this study, we measure the effectiveness and efficiency of the source search process and outcomes. The effectiveness is measured by the success rate. As the source search process can forever go on if the source is not found, we define that a successful source search process means the robot finds the actual source within 400 steps (a step means an iteration of updating search states). If the robot cannot find the source (either with or without human involvement) after 400 steps, the source search task is considered to be failed. The efficiency is measured by the number of steps the robot takes to successfully find a source. A failed source search is not taken into account in calculating the efficiency. Furthermore, we measure the average execution time per task to see how engaged the participants are during task execution.

Furthermore, we use two standard questionnaires to understand the perceived usability and cognitive workload while using the crowd-powered source searching system. The perceived usability is measured by System Usability Scale (SUS) [39]. Using the ratings of SUS items, we can derive scores of the SUS in two aspects – usability and learnability [40, 41]. Furthermore, we measure cognitive workload using NASA-task load index (NASA-TLX).

4.4   Post-Task Interview

We used SUS and NASA-TLX to measure perceived usability and cognitive workload. To have a deeper understanding of human needs, we further carried out a post-task interview/conversation with each participant. The questions used in the interview are shown in Table 1. The interview consists of 10 questions. The first and the last questions ask for general feedback and general comments. The other 8 questions are used for acquiring participants' thoughts on explainability and interaction. We asked questions 2-5 to understand the effect of explanation and how could the explanation be improved, while we asked questions 6-9 to understand the effect of interaction and how could the interaction be improved. The questions are supposed to be answered by either 7-point Likert scales (from 1 Not at all to 7 To a very large extent) or open-ended text, except question 6. Concerning question 6, participants can report

Table 1   The questions used in the post-task interview.

ID	Question	Type
1	In general, do you like this system?	7-pt Likert scale
	Could you explain why?	Open-ended
Explanability		
2	Do you understand how source searching works in this system?	7-pt Likert scale
	Could you explain why?	Open-ended
3	Do you think the user interface correctly shows how source searching works?	7-pt Likert scale
	Could you explain why?	Open-ended
4	When you were operating, was it easy to see why the machine was having a problem?	7-pt Likert scale
	Could you explain why?	Open-ended
5	Do you have any suggestions in terms of improving the explainability?	Open-ended
Interaction		
6	Which control mode do you prefer?	Multiple choices
	Could you explain why?	Open-ended
7	Do you think your ideal solution can be correctly achieved by this system?	7-pt Likert scale
	Could you explain why?	Open-ended
8	Did you successfully solve machine's problems?	7-pt Likert scale
	Could you explain why?	Open-ended
9	Do you have any suggestions in terms of improving the interaction?	Open-ended
10	Do you have any other comments?	Open-ended
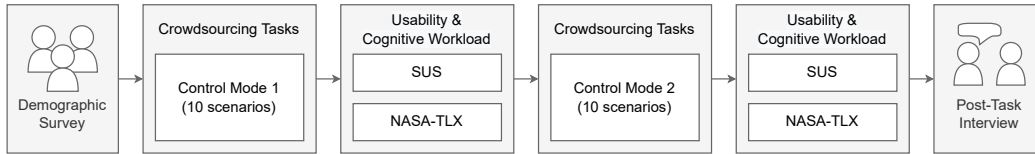
Fig. 4   The experimental procedure. A participant must complete tasks using two interaction modes (i.e., Full Control and Aided Control). However, the order of the control modes was pre-scheduled for all the participants to avoid learning bias.

what control mode they prefer (full control, aided control, both, or none).

4.5   Procedure

The experiment was organized as shown in Fig. 4. We first asked participants to complete a demographic survey. This survey requires participants to provide their basic background information about their age, gender, education level, and domain knowledge about search algorithms. After the demographic survey, we also briefly explained to the participants our experimental scenarios (i.e., to find a gas source) and how to use the prototype system.

After demographic surveys, participants were asked to complete source search crowdsourcing tasks. Each participant should complete 20 scenarios using 2 control modes, i.e., full control and aided control. To avoid learning biases, the order of the control modes during task execution was pre-scheduled – half of the participants first executed 10 full-control scenarios and then aided-control scenarios (2 experts + 3 non-experts), and the other half first executed aided-control scenarios and then full-control ones. After finishing each control mode (10 scenarios), participants were asked to rate their feelings on system usability and cognitive workload using standard questionnaires. Finally, in terms of the interview, we asked questions to the participants orally, also with printed questionnaires. The interviews were recorded, and later transcribed into text.

5   Results

We evaluated the effect of using the crowd-powered method in source search algorithms by measuring the effectiveness (success rate), the efficiency (the number of steps taken to find the source), the human execution time, the self-report SUS scores, and the self-report TLX scores.

5.1   Participants

Table 2 shows the demographic information, including gender, age, education level, expertise, and domain

Table 2   Demographic information of the participants of this study.

ID	Gender	Age	Education/Degree	Expertise	Knows search algorithms	Knows source search
1	Male	27	Master	Expert (4.5 years)	To a very large extent	To a large extent
2	Male	40	Ph.D.	Expert (2.0 years)	To a moderate extent	To a large extent
3	Male	28	Ph.D.	Expert (5.0 years)	To a very large extent	To a very large extent
4	Male	23	Bachelor	Expert (1.5 years)	To a large extent	To a large extent
5	Female	21	Undergraduate	Non-expert	To a very small extent	Not at all
6	Female	24	Bachelor	Non-expert	To a very small extent	Not at all
7	Male	30	Ph.D.	Non-expert	To a small extent	Not at all
8	Male	25	Bachelor	Non-expert	To a large extent	Not at all
9	Male	26	Master	Non-expert	To a moderate extent	Not at all
10	Female	33	Master	Non-expert	To a very small extent	Not at all

knowledge, of 10 participants recruited in our study. We asked four experts (academic researchers or engineers), who have been working on topics related to source search for at least 1 year, to participate in our study. Furthermore, we recruited 6 non-expert volunteers from our institute who had no experience in source search. People involved in the prototype system development were not invited to the experiment to avoid potential biases. The experiment was approved by the ethics committee of our institute.

5.2   Source Search Result

We evaluated source search from three perspectives, namely the effectiveness (the success rate), the efficiency (the number of steps used to find the source), and the human execution time per task. Results are shown in Table 3. Clearly, the crowd-powered method is proved to be effective, as the success rates can achieve 100% in most cases (except only one case) being approximately 22% higher than baseline 1, and 12% higher than baseline 2. It shows that leveraging human inputs could make the algorithm performance nearly perfect. Furthermore, we observe the improvement of efficiency when the full control mode is used, in comparison with both aided control and the baselines. In general, both experts and non-experts showed good performance while collaborating with the machine to solve the problems of the search algorithm.

To deeply understand the difference among experimental conditions, we performed statistical analysis for efficiency (the number of steps) and human execution time (per task). Since the numbers of search steps follow normal distributions according to the normality tests, we applied two-way ANOVA to see the effects of two factors considered in this study – expertise (expert vs non-expert) and control mode (full control vs aided control), as well as their interaction effect. Results of statistical tests are shown in Table 4. We found that the efficiency of source search shows a significant difference in terms of the control mode ($p = 0.026$), meaning the full control mode could achieve better efficiency regardless of expertise.

Since distributions of the human execution time do not come from a normal distribution according to normality tests (for all the data groups $p < 0.003$), we applied pairwise Mann-Whitney U tests to carry out the significance tests (α values were adjusted by Bonferroni corrections). We did not find a significant difference in terms of human execution time ($p > 0.07$ for all the pairs), meaning neither expertise nor control mode could significantly affect execution time.

The source search result conveys three main messages:

(1) *The crowd-powered method is effective and efficient for improving source search;*

(2) *Through our design, non-experts could achieve similar performances as experts could do;*

(3) *Taking over the machine during problem-solving could further improve the efficiency of source search.*

5.3   Usability

We asked all the participants to fill up the system usability scale (SUS) after completing each control mode (i.e., full control and aided control). Therefore, each participant provided 2 SUS responses. Since we only recruited 10 participants (20 SUS responses in total), we did not use statistical tests to perform the analysis.

Table 3 Results of the source search experiment.

Groups	Expertise	Effectiveness (% success rate)	Efficiency (# steps per task)	Human execution time (seconds per task)
Full control	Expert	100	138.85 ± 79.00	29.59 ± 25.47
	Non-expert	98	144.73 ± 87.62	34.40 ± 30.16
Aided control	Expert	100	175.10 ± 67.67	33.58 ± 27.87
	Non-expert	100	165.67 ± 80.60	29.01 ± 29.51
Baseline 1	-	78.5	154.04 ± 91.32	-
Baseline 2	-	88	179.64 ± 96.45	-

Table 4 Results of two-way ANOVA for the efficiency (# of steps) of source search.

Factors	Efficiency (# of steps)		
	dF	*F*-value	*p*-value
Expertise (Expert vs Non-expert)	1	0	0.9764
Control mode (Full control vs Aided control)	1	5.01	0.0263*
Expertise × Control mode	1	0.68	0.4089

Note: an asterisk (*) represents significant difference ($p < 0.05$).

Scores of SUS are reported in Tab.5. According to previous studies [39–41], SUS can measure the usability and learnability of a system.

While we found that the full control mode (meaning the search process is taken over by humans) showed better search efficiency, the participants in general reported that they perceived better usability and learnability from the aided control mode, rather than the full control mode. Interestingly, for all the non-experts, their SUS scores of aided control were not lower than the scores of full control, meaning they all preferred the aided control mode. However, in terms of experts, we observed more diverse opinions, and the difference of experts' overall average SUS scores between aided control and full control was less obvious (Aided Control 86.25 vs Full Control 80.63), in comparison with non-experts (Aided Control 80.42 vs Full Control 65.00).

The SUS result conveys one main message: *The participants, especially the non-experts, generally perceived better usability when they were aiding the machine, in comparison with taking over the machine.*

5.4 Cognitive Workload

To understand cognitive workload during human-machine collaboration, we used NASA-TLX scale after the participants completed each control mode (i.e., full control and aided control). Similarly, each participant also only provided 2 TLX responses, resulting in 20 TLX responses in total. To this end, we did not perform any statistical analysis. Results of NASA-TLX scores on six dimensions, namely physical demand, mental demand, temporal demand, performance, effort, and frustration, as well as the overall TLX score are reported in Tab. 6. We observed that the non-experts in general perceived less cognitive workload in the aided control mode, compared to the full control mode, across all the TLX dimensions. However, in terms of the experts, we found that all the experts thought their performances while using the full control mode were not worse (3 out of 4 reported higher) than the aided control mode. This is an opposite finding compared to non-experts. In terms of other dimensions, again, we observed more diverse opinions from the experts, resulting in tiny differences between the aided control mode and the full control mode.

The TLX result conveys two main messages:

(1) *The participants, especially the non-experts, generally perceived less cognitive workload when they were aiding the machine, in comparison with taking over the machine;*

(2) *The experts reported better performances when they took over the machine during problem-solving, while the non-experts did the opposite.*

6 Discussion

Results have shown that a crowd-powered framework could improve the effectiveness and efficiency of a source search algorithm. We also observed interesting findings in terms of usability and cognitive workload. Clearly, these findings show that experts and non-experts have different needs and preferences.

6.1 General Feedback

In general, experts liked the system very much, and non-experts also liked the system to some extent, as

Table 5 Results of the system usability. FC means the full control mode (humans take over the machine), while AC means the aided control mode (humans aid the machine).

Expertise	Participant ID	Usability	Learnability	Overall
Expert	1	AC (90.63) > FC (68.75)	AC (100.0) = FC (100.0)	AC (92.50) > FC (75.00)
	2	FC (93.75) > AC (90.63)	AC (62.50) > FC (50.00)	AC (85.00) = FC (85.00)
	3	AC (84.38) = FC (84.38)	FC (87.50) > AC (75.00)	FC (85.00) > AC (82.50)
	4	AC (81.25) > FC (71.88)	AC (100.0) = FC (100.0)	AC (85.00) > FC (77.50)
	Average	AC (86.72) > FC (79.69)	AC (84.38) = FC (84.38)	**AC (86.25) > FC (80.63)**
Non-Expert	5	AC (78.13) > FC (59.38)	AC (87.50) > FC (50.00)	AC (80.00) > FC (57.50)
	6	AC (87.50) > FC (68.75)	AC (87.50) > FC (62.50)	AC (87.50) > FC (67.50)
	7	AC (87.50) > FC (84.38)	AC (100.0) = FC (100.0)	AC (90.00) > FC (87.50)
	8	AC (71.88) > FC (65.63)	AC (100.0) > FC (87.50)	AC (77.50) > FC (70.00)
	9	AC (87.50) > FC (71.88)	AC (37.50) = FC (37.50)	AC (77.50) > FC (65.00)
	10	AC (78.13) > FC (43.75)	AC (37.50) = FC (37.50)	AC (70.00) > FC (42.50)
	Average	AC (81.77) > FC (65.63)	AC (75.00) > FC (62.50)	**AC (80.42) > FC (65.00)**

Table 6 Results of the cognitive workload. FC means the full control mode (humans take over the machine), while AC means the aided control mode (humans aid the machine).

Expertise	ID	Physical demand	Mental demand	Temporal demand	Performance	Effort	Frustration	Overall TLX
Expert	1	AC (0) < FC (5)	AC (0) < FC (5)	AC (5) < FC (20)	AC (0) = FC (0)	AC (0) < FC (15)	AC (0) = FC (0)	AC (0.8) < FC (7.5)
	2	FC (65) < AC (80)	FC (50) < AC (70)	FC (0) < AC (5)	FC (15) < AC (50)	FC (50) < AC (70)	AC (50) = FC (50)	FC (38.3) < AC (54.2)
	3	AC (10) < FC (20)	AC (5) < FC (30)	AC (0) = FC (0)	FC (0) < AC (10)	AC (10) < FC (15)	AC (0) = FC (0)	AC (5.8) < FC (10.8)
	4	AC (60) = FC (60)	AC (60) < FC (80)	AC (0) = FC (0)	FC (0) < AC (5)	FC (35) < AC (40)	AC (40) < FC (75)	AC (34.2) < FC (41.7)
	Average	AC (37.5) = FC (37.5)	AC (33.8) < FC (41.3)	AC (2.5) < FC (5.0)	FC (3.8) < AC (16.3)	FC (28.8) < AC (30.0)	AC (22.5) < FC (31.3)	**AC (23.8) < FC (24.6)**
Non-Expert	5	AC (5) = FC (5)	AC (0) < FC (5)	AC (0) = FC (0)	AC (5) < FC (50)	AC (10) < FC (15)	FC (5) < AC (10)	AC (5.0) < FC (13.3)
	6	AC (5) < FC (35)	AC (5) < FC (10)	AC (10) < FC (25)	AC (0) < FC (10)	AC (15) < FC (20)	FC (5) < AC (10)	AC (7.5) < FC (17.5)
	7	FC (15) < AC (20)	AC (0) = FC (0)	AC (0) < FC (20)	AC (50) = FC (50)	FC (30)< AC (35)	AC (60) = FC (60)	AC (27.5) < FC (29.2)
	8	AC (0) < FC (10)	AC (0) < FC (10)	AC (0) = FC (0)	AC (0) = FC (0)	AC (0) < FC (50)	AC (0) < FC (50)	AC (0.0) < FC (20.0)
	9	AC (0) = FC (0)	AC (0) < FC (50)	FC (15) < AC (20)	FC (15) < AC (20)	AC (25) < FC (50)	AC (0) < FC (20)	AC (9.2) < FC (24.2)
	10	AC (5) < FC (10)	AC (0) = FC (0)	AC (0) < FC (15)	FC (15) < AC (25)	FC (0) < AC (10)	AC (5) < FC (15)	AC (7.5) < FC (9.2)
	Average	AC (5.8) < FC (12.5)	AC (0.8) < FC (12.5)	AC (5.0) < FC (12.5)	AC (15.0) < FC (22.5)	AC (15.8) < FC (27.5)	AC (14.2) < FC (25.8)	**AC (9.4) < FC (18.9)**

shown in Fig. 5 (a). Experts thought that the system was professional, since it could "present various kinds of important factors and information for source searching in a professional way" (Participant 4, Male, Age 23, Expert), and it could effectively address practical problems. Particularly, the experts believed that humans play a very important role here, since the machine was not always reliable and the AI's suggestions of estimated source locations were not always precise.

According to experts' understanding, during source searching, there is a trade-off between immediately going to the AI-suggested source location and exploring other areas to gain more information (we also observed that non-experts would like to immediately guide the robot to the suggested location of the source, without extra exploration). Humans were there to manage this subtle trade-off (Participant 1 and Participant 2).

Another important factor is trust. When the AI was explaining the search algorithm to the participants, some participants chose to not trust the AI (Participant 8). However, some participants trusted it very much at the beginning but felt deceived when the AI was wrong

(Participant 9). After reading participants' feedback, we argue that a further explanation is needed to show the confidence level of the AI, so that users will understand to what extent they should trust AI, as studied by previous work as well [10, 16, 17].

In summary, the participants in general liked the system. Both experimental results and participants' feedback show that humans could play a critical role in the search process. While people consider human-AI collaboration as a professional and effective means, an important issue has been raised concerning human trust in AI systems. Trust is a key driver of inter-human collaboration. As AI techniques advance, building trust would be a big leap towards future human-AI collaboration.

6.2 How Can AI Explanation and Suggestion Be Improved?

In the post-task interview, we explicitly asked participants questions about their feeling on the current explanation & suggestion method we used, and how the method could be improved. As can be seen in Fig. 5 (b), (c), and (d), experts generally appreciated the expla-
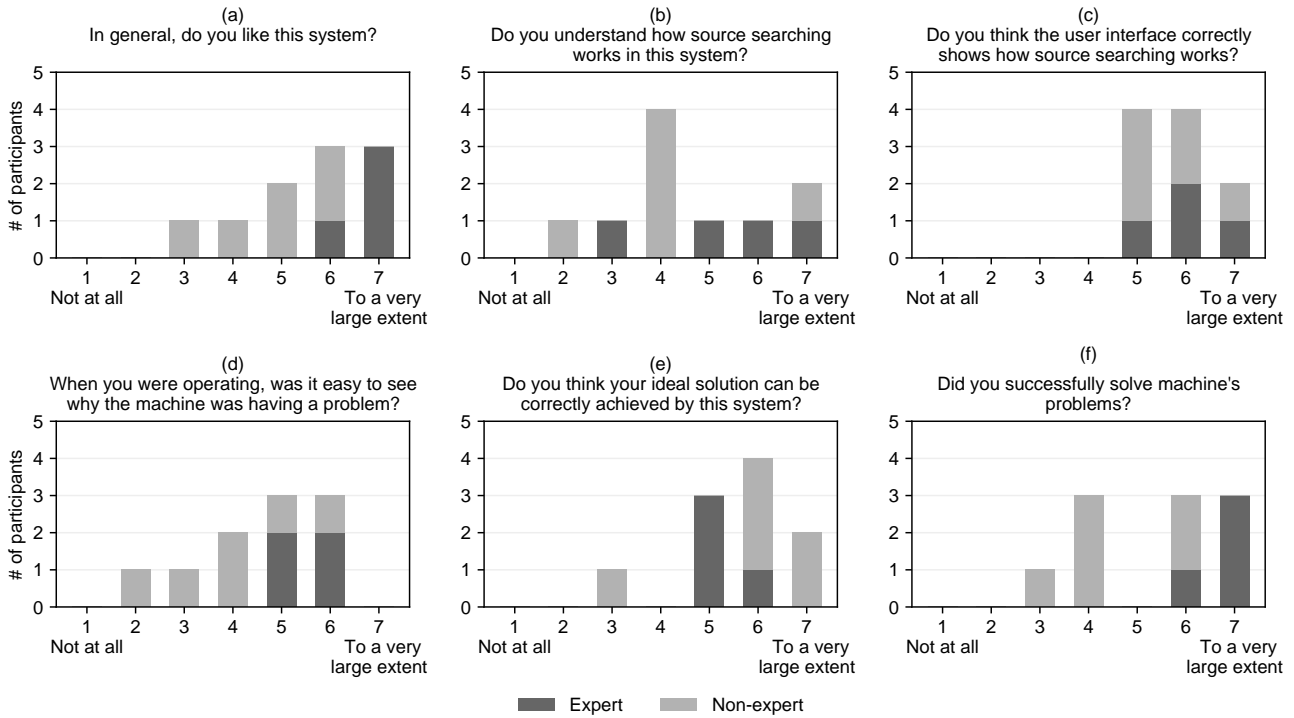
Fig. 5   Histograms of the questions that are answered by 7-point Likert scales.

nation and understood the problem well, whereas non-experts did not report good scores in explanation. We consider this finding reasonable since the prototype system was designed according to experts' discussions and only from experts' perspectives.

After carefully analyzing all the comments from 10 participants, we suggested that the search algorithm explanation should particularly look into two aspects: **personalized explanation** and **explanation in natural language**.

6.2.1   Personalized Explanation

In the prototype system, the user interface used graphical elements with some text to explain the search algorithm as well as the problem the algorithm encountered. The way of explanation was identical to both experts and non-experts. Participants' feedback suggested that personalization would be helpful in explanation. In general, experts reported that they looked forward to more information (2 out of 4), while non-experts reported that they looked forward to less information (3 out of 6). In addition, it is interesting to see that there were no experts wanted less information and no non-experts wanted more information.

We could learn from the feedback that the focus of people differed a lot when they tried to understand the algorithm and the problem, meaning the explanation should be customizable and able to satisfy various preferences. More information for experts and less information for non-experts is just one direction that personalized explanation could advance forward. As previous studies have shown that a variety of factors such as human emotions and moods could affect the performance [42–46], we argue that studying personalization in many different aspects is a crucial topic in developing effective and friendly human-AI collaborative systems.

6.2.2   Explanation in Natural Language

Using natural language and conversation as the media for human-AI interaction has become a trend. Previous works have shown the advantages of using conversational interfaces over traditional graphical interfaces, for improving satisfaction, engagement, and output quality [47–49]. In terms of human-AI collaboration for search algorithms, we found that the participants would like to see a more natural communication means – instead of giving professional explanations using graphical elements, both experts and non-experts looked forward to explanations in natural language, which could either be text-based or voice-based.

Future work could focus on developing novel methods for automatically generating understandable explanations in human languages. As the current prototype system and explanation methods used in this work did

not help non-experts much in understanding the algorithm and the problem, it is worth exploring whether explanation in natural language could further improve the user experience for non-experts.

6.3 How Can Crowd-Powered Problem Solving Be Improved?

To understand whether the current crowd-powered problem-solving approach meets participants' needs, we explicitly asked questions in the post-task interview. In terms of the question "which control mode do you prefer?", only one participant (expert) reported that he preferred the full control mode, meaning humans took over the machine during problem-solving. All the other 9 participants preferred the aided control mode, meaning humans aiding the machine during problem-solving. Furthermore, as shown in Figure 5 (e) and (f), it is interesting that non-experts tended to think that their ideal solution was correctly achieved by the system, while experts gave moderate scores. This is possibly due to the fact that experts tended to have higher standards and expected optimal solutions. However, experts believed that eventually the problems were solved successfully by them. In the following sub-sections, we analyzed each participant's comments in terms of crowd-powered problem-solving, and provided insights for improving search algorithms using human-AI collaboration from two perspectives: **personalized interaction** and **intuitive control**.

6.3.1 Personalized Interaction

We have learned that participants would like to see personalized AI explanations. In terms of interaction and problem-solving, we could learn similar findings from participants' feedback. First of all, the control mode (full control vs aided control) could be personalized. It is interesting to see that most participants preferred the aided control mode, although full control could achieve better efficiency. The main reason that most participants preferred the aided control mode is that they could put less effort but it could achieve the same goal as the full control mode could do.

Both experts and non-experts could see the advantages of using a full control mode, though most of them preferred the mode of aided control. We still argue that the full control mode could have better effects for some specific scenarios which require more precise control and more explorations, like Participant 2 (Male, Age 40, Expert) explained: "I think the environments in the current system are relatively simple, so the aided control

mode is better; when the environment becomes more complex, a full control mode might play a more important role".

6.3.2 Intuitive Control

In our context, we create a term – "intuitive control" – to represent a problem-solving approach that enables humans to make a vague goal (which does not define specific actions) for the search algorithm, following their instincts. As Participant 10 said, "Sometimes I just wanted to guide the robot to search in a certain direction rather than to move to a specific position". For the AI system, this requires a better capability of understanding human intention, and therefore it can further reduce human cognitive workload and achieves a more intelligent collaboration.

6.4 Implications for Designing Crowd-Powered Systems

Our study has shown that using human-AI collaboration is effective in improving source search algorithms. We have learned important lessons from the experimental result in terms of how to better leverage human intelligence. The findings provide important implications in terms of designing general crowd-powered systems.

Personalization. According to the participants' feedback, personalization is one of the most important factors that should be considered during the design phase. Our experiment shows that the design could achieve good efficiency and effectiveness, but did not satisfy non-experts' needs to some extent. Therefore, we suggest that personalization should play an important role throughout the entire process, since human needs vary a lot due to different backgrounds, education levels, and personalities.

Learning from Humans. We argue that a crowd-powered system in the future should be able to learn human behavior and accordingly adjust its own problematic actions. The prototype system has successfully improved the effectiveness and efficiency of the state-of-the-art search algorithms. However, our proposed approach still frames the capability of the machine, meaning the machine could only detect the problem and explain the problem in a way that the task designers decided, and the humans could only help the machine in problem-solving using pre-defined control modes. Human-AI collaboration should be mutually beneficial, as machines could help humans in problem-solving, and simultaneously learn from humans.

Crowd Computing. The experiment has shown that non-experts could achieve comparable output quality, with regard to effectiveness, efficiency, and execution time, in comparison with experts. This is a rather positive finding, implying the potential of massive deployment using crowd computing techniques and the possibility of leveraging swarm intelligence. Researchers and practitioners in the field of human-AI collaboration should focus on lowering the barrier of interaction, to access a larger number of users. Crowd computing has provided numerous new opportunities for multiple disciplines, including AI and HCI communities [50]. It could be connected to prevalent crowdsourcing platforms to access more diverse users, and to acquire faster responses.

6.5 Limitations and Future Work

In this study, we only recruited participants from our institute, and all the participants were either full-time researchers or students. We acknowledge the limitation that the participants in our study are not representative enough. Future work could properly perform a power analysis, consider a larger sample size, and probably recruit participants from online freelancing/crowdsourcing marketplaces, to obtain more general findings.

We also realized that the problem detection and explanation are rather simple in the prototype system design, and only two control modes were implemented in the prototype system. We consider it reasonable since this is a first step studying crowd-powered methods and human-AI collabortion in improving source search algorithms. As results indeed showed the feasibility of using human-AI collaboration, we suggest that future work could use more recent and advanced AI-based techniques to achieve a more intelligent interaction.

7 Conclusions

In this work, we proposed a research question to investigate the effects of using crowd-powered approaches in source search algorithms. To answer the research question, we designed a framework enabling human-AI collaboration for improving existing source search algorithms and carried out an experiment, asking experts and non-experts to complete 200 source search scenarios (704 crowdsourcing tasks generated in total). The experimental results showed the feasibility of crowd-powered source search in improving both effectiveness and efficiency. Furthermore, we explicitly asked the

participants to report their perceived usability, cognitive workload, and their feedback to deeply understand their needs. Finally, we provided design implications for future human-AI collaboration and crowd computing research.

Acknowledgment

References

[1] B. Ristic, A. Skvortsov, and A. Gunatilaka, "A study of cognitive strategies for an autonomous search," *Information Fusion*, vol. 28, pp. 1–9, 2016.

[2] M. Hutchinson, H. Oh, and W.-H. Chen, "Entrotaxis as a strategy for autonomous search and source reconstruction in turbulent conditions," *Information Fusion*, vol. 42, pp. 179–189, 2018.

[3] S. Desale, A. Rasool, S. Andhale, and P. Rane, "Heuristic and meta-heuristic algorithms and their relevance to the real world: a survey," *Int. J. Comput. Eng. Res. Trends*, vol. 351, no. 5, pp. 2349–7084, 2015.

[4] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen *et al.*, "Guidelines for human-AI interaction," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[5] M. J. Franklin, D. Kossmann, T. Kraska, S. Ramesh, and R. Xin, "CrowdDB: answering queries with crowdsourcing," in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, 2011, pp. 61–72.

[6] A. Bozzon, M. Brambilla, and S. Ceri, "Answering search queries with crowdsearcher," in *Proceedings of the 21st international conference on World Wide Web*, 2012, pp. 1009–1018.

[7] L. Bradeško, M. Witbrock, J. Starc, Z. Herga, M. Grobelnik, and D. Mladenić, "Curious Cat–mobile, context-aware conversational crowdsourcing knowledge acquisition," *ACM Transactions on Information Systems (TOIS)*, vol. 35, no. 4, pp. 1–46, 2017.

[8] T.-H. K. Huang, J. C. Chang, and J. P. Bigham, "Evorus: A crowd-powered conversational assistant built to automate itself over time," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. 295.

[9] I. Arous, L. Dolamic, J. Yang, A. Bhardwaj, G. Cuccu, and P. Cudré-Mauroux, "Marta: Leveraging human rationales for explainable text classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 7, 2021, pp. 5868–5876.

[10] Y. Zhang, Q. V. Liao, and R. K. Bellamy, "Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 295–305.

[11] Q. V. Liao, D. Gruen, and S. Miller, "Questioning the AI: informing design practices for explainable ai user experiences," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–15.

[12] L. Von Ahn and L. Dabbish, "Designing games with a purpose," *Communications of the ACM*, vol. 51, no. 8, pp. 58–67, 2008.

[13] J. Yang, T. Drake, A. Damianou, and Y. Maarek, "Leveraging crowdsourcing data for deep active learning an application: Learning intents in alexa," in *Proceedings of the 2018 World Wide Web Conference*, 2018, pp. 23–32.

[14] S. Qiu, A. Psyllidis, A. Bozzon, and G.-J. Houben, "Crowd-mapping urban objects from street-level imagery," in *The World Wide Web Conference*, 2019, pp. 1521–1531.

[15] S. Sharifi Noorian, S. Qiu, A. Psyllidis, A. Bozzon, and G.-J. Houben, "Detecting, classifying, and mapping retail storefronts using street-level imagery," in *Proceedings of the 2020 International Conference on Multimedia Retrieval*, 2020, pp. 495–501.

[16] B. Green and Y. Chen, "Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments," in *Proceedings of the conference on fairness, accountability, and transparency*, 2019, pp. 90–99.

[17] ——, "The principles and limits of algorithm-in-the-loop decision making," vol. 3, no. CSCW. ACM New York, NY, USA, 2019, pp. 1–24.

[18] M. A. Goodrich and A. C. Schultz, *Human-robot interaction: a survey*. Now Publishers Inc, 2008.

[19] A. Kolling, P. Walker, N. Chakraborty, K. Sycara, and M. Lewis, "Human interaction with robot swarms: A survey," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 9–26, 2015.

[20] S. Bashyal and G. K. Venayagamoorthy, "Human swarm interaction for radiation source search and localization," in *2008 IEEE Swarm Intelligence Symposium*. IEEE, 2008, pp. 1–8.

[21] D. J. Bruemmer, D. D. Dudenhoeffer, M. D. McKay, and M. O. Anderson, "A robotic swarm for spill finding and perimeter formation," IDAHO NATIONAL ENGINEERING AND ENVIRONMENTAL LAB IDAHO FALLS, Tech. Rep., 2002.

[22] J. Casper and R. R. Murphy, "Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 33, no. 3, pp. 367–385, 2003.

[23] X.-x. Chen and J. Huang, "Odor source localization algorithms on mobile robots: A review and future outlook," *Robotics and Autonomous Systems*, vol. 112, pp. 123–136, 2019.

[24] T. Jing, Q.-H. Meng, and H. Ishida, "Recent progress and trend of robot odor source localization," *IEEJ Transactions on Electrical and Electronic Engineering*, 2021.

[25] Y. Zhao, B. Chen, X. Wang, Z. Zhu, Y. Wang, G. Cheng, R. Wang, R. Wang, M. He, and Y. Liu, "A deep reinforcement learning based searching method for source localization," *Information Sciences*, vol. 588, pp. 67–81, 2022.

[26] N. E. Humphries, N. Queiroz, J. R. Dyer, N. G. Pade, M. K. Musyl, K. M. Schaefer, D. W. Fuller, J. M. Brunnschweiler, T. K. Doyle, J. D. Houghton *et al.*, "Environmental context explains lévy and brownian movement patterns of marine predators," *Nature*, vol. 465, no. 7301, pp. 1066–1069, 2010.

[27] L. Champagne, R. G. Carl, and R. Hill, "Agent models II: Search theory, agent-based simulation, and u-boats in the bay of biscay," in *Proceedings of the 35th conference on Winter simulation: driving innovation*. Citeseer, 2003, pp. 991–998.

[28] A. R. Unwin, "Search theory and applications," *Journal of the Operational Research Society*, vol. 33, no. 5, pp. 494–495, 1982.

[29] Y. Zhao, B. Chen, Z. Zhu, F. Chen, Y. Wang, and D. Ma, "Entrotaxis-jump as a hybrid search algorithm for seeking an unknown emission source in a large-scale area with road network constraint," *Expert Systems with Applications*, p. 113484, 2020.

[30] S. Pang and J. A. Farrell, "Chemical plume source localization," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, no. 5, pp. 1068–1080, 2006.

[31] M. Hutchinson, H. Oh, and W.-H. Chen, "A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors," *Information Fusion*, vol. 36, pp. 130–148, 2017.

[32] A. Marjovi, J. G. Nunes, L. Marques, and A. De Almeida, "Multi-robot exploration and fire searching," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 1929–1934.

[33] M. Dadgar, S. Jafari, and A. Hamzeh, "A PSO-based multi-robot cooperation method for target searching in unknown environments," *Neurocomputing*, vol. 177, pp. 62–74, 2016.

[34] H. Tang, W. Sun, H. Yu, A. Lin, and M. Xue, "A multirobot target searching method based on bat algorithm in unknown environments," *Expert Systems with Applications*, vol. 141, p. 112945, 2020.

[35] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan, "An introduction to mcmc for machine learning," *Machine learning*, vol. 50, no. 1, pp. 5–43, 2003.

[36] A. Smith, *Sequential Monte Carlo methods in practice*. Springer Science & Business Media, 2013.

[37] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN revisited, revisited: why and how you should (still) use DBSCAN," *ACM Transactions on Database Systems (TODS)*, vol. 42, no. 3, pp. 1–21, 2017.

[38] M. Vergassola, E. Villermaux, and B. I. Shraiman, "'Infotaxis' as a strategy for searching without gradients," *Nature*, vol. 445, no. 7126, pp. 406–409, 2007.

[39] J. Brooke *et al.*, "SUS-A quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[40] J. R. Lewis and J. Sauro, "The factor structure of the system usability scale," in *International conference on human centered design*. Springer, 2009, pp. 94–103.

[41] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *Intl. Journal of Human–Computer Interaction*, vol. 24, no. 6, pp. 574–594, 2008.

[42] T. A. Wright and R. Cropanzano, "The happy/productive worker thesis revisited." 2007.

[43] P. M. Desmet, M. H. Vastenburg, and N. Romero, "Mood measurement with pick-a-mood: review of current methods and design of a pictorial self-report scale," *Journal of Design Research*, vol. 14, no. 3, pp. 241–279, 2016.

[44] C. Beedie, P. Terry, and A. Lane, "Distinctions between emotion and mood," *Cognition & Emotion*, vol. 19, no. 6, pp. 847–878, 2005.

[45] P. Verduyn, I. Van Mechelen, and F. Tuerlinckx, "The relation between event processing and the duration of emotional experience." *Emotion*, vol. 11, no. 1, p. 20, 2011.

[46] M. Zhuang and U. Gadiraju, "In what mood are you today? an analysis of crowd workers' mood, performance and engagement," in *Proceedings of the 10th ACM Conference on Web Science*, 2019, pp. 373–382.

[47] R. J. Moore, R. Arar, G.-J. Ren, and M. H. Szymanski, "Conversational UX design," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2017, pp. 492–497.

[48] S. Kim, J. Lee, and G. Gweon, "Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality," in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–12.

[49] S. Qiu, U. Gadiraju, and A. Bozzon, "Improving worker engagement through conversational microtask crowdsourcing," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–12.

[50] U. Gadiraju and J. Yang, "What can crowd computing do for the next generation of ai systems?" in *CSW@ NeurIPS*, 2020, pp. 7–13.

Yong Zhao received the B.E. and M.E degree in Simulation Engineering and Control Science and Engineering from National University of Defense Technology, China, in 2019 and 2021, where he is currently pursuing the Ph.D. degree with Mangement Science and Engineering. His research interests include the parallel systems method, autonomous searching and crowd sensing. He is now working on localization the odor source using mobile sensors.

Zhengqiu Zhu received the B.E. and M.E. degree in Simulation Engineering from National University of Defense Technology, Changsha, China, in 2016 and 2018, where he is currently pursuing the Ph.D. degree in the College of Systems Engineering. He was also a visiting Ph.D. student in the research group of Multi-scale Networked Systems at University of Amsterdam (UvA). He has authored or coauthored more than 30 publications in international refereed journals and conferences. His research interests include mobile crowdsensing, ubiquitous computing, and social computing.

Bin Chen received his B.E., M.E. and Ph.D. degree in Control Science and Engineering from National University of Defense Technology in 2003, 2005 and 2010. He is currently an associate research fellow in the College of Systems Engineering at National University of Defense Technology. He has authored or coauthored more than 100 publications in international refereed journals and conferences. His current research focuses on ACP-based simulation running support, parallel experimental method, data mining, mobile crowdsensing and micro-task crowdsourcing.

Sihang Qiu received his Ph.D. from the Web Information Systems group of TU Delft in 2021 and is an associate professor at College of Systems Engineering of National University of Defense Technology. He is interested in developing novel human-AI interaction techniques. His research also focuses on human-centered AI, crowd computing, and conversational agents.