

Follow the Yellow or Red Brick Road? Investigating the Impact of Narratives in a Guided Navigation Task

Anonymous Author(s)

ABSTRACT

Sharing stories during casual conversations is common and inherently human, adding personal and unique details to what we say. Including them in dialogue provides measurable advantages such as higher memory recall and more positive social human-human interactions. However, with the increasing presence of social robots in everyday environments, whether these stories, or narratives, have similar effects on human-robot interactions with embodied agents is unclear.

In this paper, we develop and perform a pilot evaluation with a direction-giver agent to study the effects of narratives in the interaction. We use a Furhat robot with an LLM to guide participants in a video game task. Along with detailed instructions, the robot answers follow-up questions at any time during the route-following task, engaging in short conversations about what to do conditioned on the participant's actions. We assess the feasibility of using a virtual environment for these tasks which traditionally would require a physical environment, and present design guidelines for their use. We measure the impact that the use of narratives has in the interaction, such as with robot social acceptability scales, and find early indications that participants made fewer navigational mistakes and preferred narratives in social acceptability measures. We end with a discussion on a planned larger-scale experiment and the limitations of our current study.

CCS CONCEPTS

- Human-centered computing → Interactive systems and tools; • Computer systems organization → Robotics.

KEYWORDS

Social agents, narratives, navigation, design guidelines

ACM Reference Format:

Anonymous Author(s). 2024. Follow the Yellow or Red Brick Road? Investigating the Impact of Narratives in a Guided Navigation Task. In *Proceedings of The 12th International Conference on Human-Agent Interaction (HAI '24)*. ACM, New York, NY, USA, 9 pages. <https://doi.org/XXXXXX.XXXXXXX>

1 INTRODUCTION

When navigating an unknown environment, or finding their way along a route, a person may use various spatial, cognitive, and behavioural abilities [35]. While it varies from person to person,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

HAI '24, November 24–27, 2024, Swansea, UK

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/18/06

<https://doi.org/XXXXXX.XXXXXXX>



Figure 1: A user interacting with the experimental Furhat setup.

aspects of navigation can be made “easier”, where one such avenue is to use landmarks, which have been suggested to improve the navigational efficiency and reliability of route instructions [11]. An alternative strategy involves adding narrative elements to the instruction set, potentially fundamental to comprehending our everyday reality [5]. In our task environment, we aim to use a narrative’s ability to increase social acceptance of the robot, alongside usefulness and recall of the instruction set [36]. Humans commonly perform this “story-telling in directions” as a way to aid their recall. For instance, a study by Hook et al. [20] documented how tourists without prior knowledge of the environment received route directions. Notably, the route describers added small stories to important buildings or landmarks to facilitate recall. However, whether these narrative effects can be exploited in a Human-Agent Interaction (HAI) perspective remains an open question, especially when humans are trying to complete a task with the aid of a narrative robot. Social robots that can aid humans with tasks are already appearing in various public and private spaces, such as baristas [25] or homes [10], however, little research has been performed on the usefulness of narratives in HAI.

In this paper, we present the development of a Large Language Model (LLM)-based direction-giving robot and a subsequent pilot evaluation in a lab setting, shown in Figure 1. We used a Furhat robot¹ to guide participants in a virtual environment, Minecraft², in two conditions: narrative or non-narrative. Throughout the navigation task, participants can ask follow-up questions to the robot to clarify their next action or obtain reminders of the navigational steps.

In this paper, we aim to assess the feasibility and potential advantages of using a virtual environment for our assessment rather than a physical environment. We contribute a set of design guidelines based on this assessment to help future studies that hope to

¹<https://furhatrobotics.com/>, last accessed: 20/05/2024

²<https://www.minecraft.net/>, last accessed: 20/05/2024

complete navigation tasks within modifiable virtual environments such as Minecraft. As part of our pilot evaluation, we also study several social robot acceptance factors in preparation for a larger study, namely: perceived usefulness, perceived likeability, perceived animacy, alongside the impact on recall and the number or type of follow-up questions, depending on the conversation style used.

2 RELATED WORK

Narratives are highly interdisciplinary in their nature, with huge quantities of research existing across multiple fields in an attempt to quantify their core formation, their effect on behaviour, and why this effect exists [16].

Definitions across disciplines vary [26, 31, 33, 37], however, the fundamental principal requires that narratives are built from a series of coherent events, linked with causal, temporal, and spatial chains. Each of these chains furthers cohesiveness in the narrative, but each chain is not always present. To demonstrate the effect of coherence, an example is drawn from Per Persson et al. [34]:

"A woman is dreaming about cars. A window breaks. A profit is being made."

Based on the above text, it can be difficult to relate each event in any sort of causal, temporal, or spatial way. The connections are vague and indeterminate, and if a reader was trying to interpret any sort of story, they may assume no coherence between the statements. However, if one considers the aforementioned chains:

"Sandra had dreamed about a new car for several months and last week she bought one down town. The next morning as she was going to work, the car wouldn't start. She got so angry she smashed the window of the car and her hand started to bleed. Next week she sold the car - with a profit."

Now, a much more coherent narrative is created. Each event is related spatially and temporally, with references to time and indications that the car had been taken home. More importantly, each event can be related causally, with the three main narrative events being a series of cause and effects. The dream car causes the purchase; the fact that it is broken causes the anger and in response the damage to the window and her hand; and finally the prior events cause her to reconsider and sell the car. Coherence is thus achieved when elements in a text can be related using these three chains [34].

Coherent narratives are utilised in a wide variety of fields to exploit effects. For example, in medicine, narratives have been used to demonstrate experiences of others as a decision aid, able to provide information, convey empathy, and change behaviour [3, 37]. Likewise, in journalism, narratives are used in the form of exemplars, where the experiences of individuals are used to give life to stories or explain abstract concepts [19].

Both of these examples point towards the use of narratives to improve recall, and improve communication effectiveness through engagement. In the context of recall, Kilaru et al. [23] and Graesser et al. [14] investigate narrative text versus factual text, in the form of physician guidelines and encyclopedic text respectively. Both studies observed a higher short-term recall from the narrative text, with the authors additionally observing an approximately twice as fast reading rate. This pattern was backed up further by Graesser

et al. [15], again demonstrating a faster reading time with narrative text.

With regards to communication effectiveness, narratives have exhibited more engaging behaviour than didactic presentations of information. Volk et al. [38] evaluated an entertainment-based decision aid for prostate cancer screening, splitting participants into low and high literacy categories. Comparing against an audiobooklet-control aid, increased engagement was reported by participants at a low-literacy level when using the entertainment-based decision aid, with a particularly high engagement reported with celebrity based narrative testimonials. Similar work by Jibaja-Weiss et al. [22] supported this work, with a entertainment education system for early stage breast cancer surgery choices. In their study to inform lower health-literate woman about breast cancer surgery options, the narrative strategy was found to be a more desirable strategy than non-narrative. It was also reported that the narrative messaging was more engaging than non-narrative as the information conveyed is perceived to be more relevant to the recipient.

Way-finding here can be described as purposeful, motivated movement from a starting point to a specific destination, which is not able to be directly perceived by the navigator, and taking place along a route [13]. Route descriptions are ordinarily split into several steps which can then be individually verbalised [9]. Each instruction step can be split into two components, contributing functions in the conversation discourse that must be viewed separately [1]. Firstly, an action that the navigator can perform to change the environment, e.g. "turn", "walk". Secondly, a environmental descriptor of where the action should be performed, e.g. "next to the red car". The "red car" in the example above is acting as a landmark, which play a crucial role in communicating route directions [6]. By providing references to the environment, identification of critical decision making points is made easier, and navigators can construct a mental representation of an unfamiliar environment easier [30]. Clarity of specific route instructions is also improved with landmarks, improving navigation efficiency and reliability [11].

Researchers have also looked at this issue from a human-robot perspective. Heikkila et al. [18] investigate a route guidance task from a HAI perspective. A robot is situated in a shopping center, with one of its abilities being to give humans route guidance. After conducting a four-phased qualitative study, they found nine design implications, one of which noted that salient landmarks and those located in the crossings of aisles are helpful, but one must moderate their use.

Where narratives could be used to further improve the abilities of landmarks in recall and usefulness is lacking research. Work by [34] touches on the potential for narratives in digital navigation. In this work, points are made referring to the impact of the narrative on the navigator, first noting that if a navigator learns fast and remembers well, they will be better equipped to deal with dead ends and make shortcuts. Second, it is noted that the narrative mode may make the navigator remember the path better, however it may also increase their ability to describe the path to others more than other semantic-abstract modes. Third, they note that narrative appears to be a natural mode which people use to better memorise spatial descriptions, and as such it would be natural for their ability to memorise and generate descriptions of routes for navigational

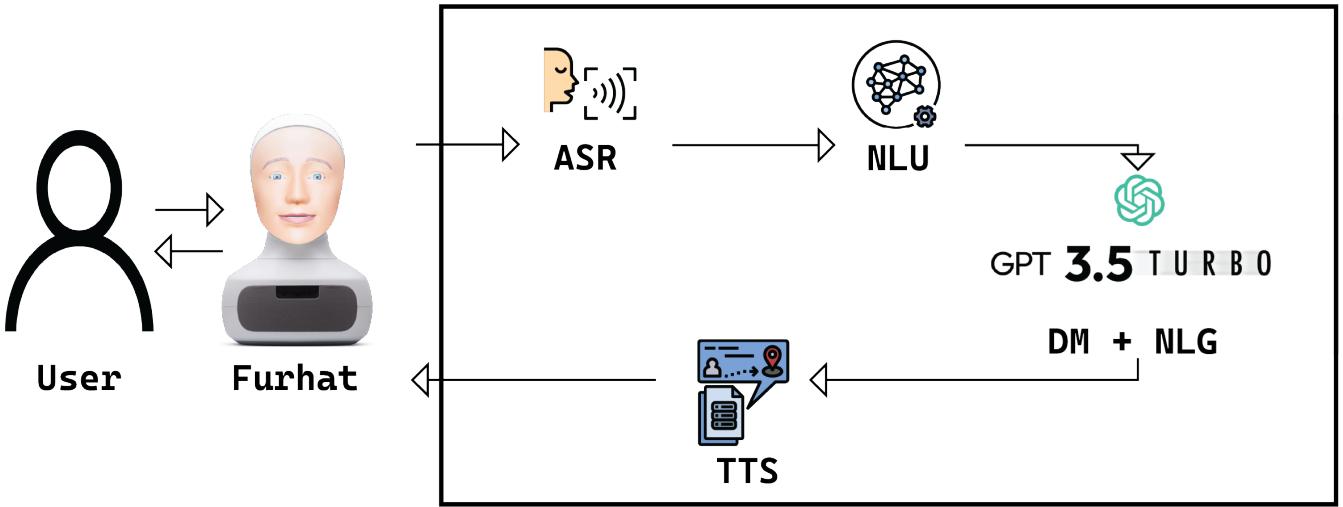


Figure 2: System architecture diagram showing the connected components. Icons made by noomtah from flaticon.com. NLG (Natural Language Generation).

purposes to increase. Finally, they note that if narratives are used at the point of landmarks, then the narratives must remain stable and unchanging much like the landmarks in order for the navigator to recognise the narrative if mentioned again.

3 SYSTEM ARCHITECTURE

The system consists of a Furhat robot connected to an LLM and the virtual environment. Users can interact with Furhat through speech and use a monitor, mouse, and keyboard attached to a laptop to move in the environment, see Figure 3.

3.1 Embodied Agent

We use a Furhat as the embodied robot in our setup. It has the form of a human-sized head with a hat and a back-projected face that allows a wide range of faces and gestures. Its social capabilities go beyond text-to-speech, providing natural facial expressions and

non-verbal cues to match the text, as well as voice or tone changes. The Software Development Kit (SDK)³ has tools to manage the interaction flow with policies such as user recognition and gaze-following.

Regarding its speech capabilities, we selected a high-quality and clear voice as instructions would sometimes be long and users would be in unfamiliar environments. We use external software to handle Automatic Speech Recognition (ASR) and Text-To-Speech (TTS): Google Cloud Speech-To-Text⁴ and Amazon Polly⁵ respectively. We also use an attached high-quality microphone to complement Furhat's microphones. Similar to smart speakers or voice interfaces, a light underneath the robot lights up to signal it is speaking (blue) or listening to the user (green).

We use Furhat's native Natural Language Understanding (NLU) and Dialogue Management (DM) for the starting interaction. To avoid introducing unexpected events, we handcrafted an initial conversation flow restricted to defined paths built with custom intents on top of the NLU model. This includes the initial greeting, some chit-chat between the user and the Furhat and providing instructions for the environment. After the user confirms that they understand the instructions and during the interaction with the environment, the system uses an LLM as the DM/NLG instead. Figure 2 shows the full system architecture.

3.2 LLM

We use a Large Language Model (LLM) at the core of the NLP capabilities in our system. LLMs have shown impressive capabilities in many tasks [4, 8, 39], largely removing the need for fine-tuning domain-specific models and considerably speeding up testing and deployment of conversational agents. We exploit the LLM's ability to adapt to the input text in zero-shot settings [24, 32], similar to

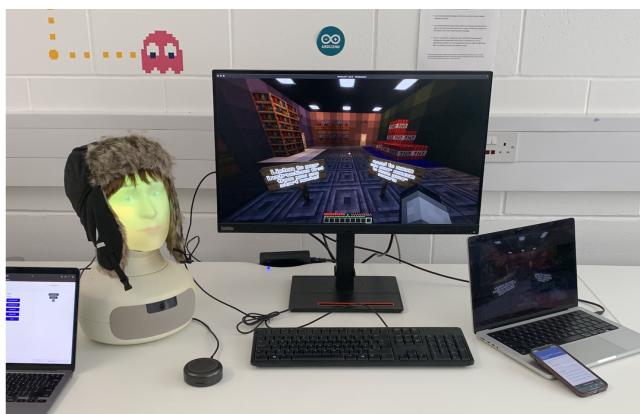


Figure 3: The experimental setup with Furhat and virtual environment.

³https://docs.furhat.io/getting_started/, last accessed: 20/05/2024

⁴<https://cloud.google.com/speech-to-text>, last accessed: 20/05/2024

⁵<https://aws.amazon.com/polly/>, last accessed: 20/05/2024

349 previous works that have explored using LLMs as the central part
 350 of a system or robotic agent [7, 21].

351 Specifically, we employ GPT-3.5 Turbo `text-davinci-003`⁶ to
 352 handle follow-up questions from users and maintain a conversation.
 353 The model can respond to clarification questions regarding the
 354 instructions or the environment (i.e., “*Should I go through the red or*
 355 *the blue door?*”) at any time and maintain a simple conversational
 356 context.

357 We query the model using the OpenAI API⁷, which often in-
 358 curred a delay. To maintain the flow of the interaction, we introduce
 359 backchannels or short utterances that Furhat speaks out loud if
 360 the API response is not back after a threshold⁸ (i.e., “*Hmmm...*” or
 361 “*Okay, let’s see*”), such as the dialogue in Figure 4.

362 **3.2.1 Prompting.** We elicit suitable responses from the LLM using
 363 prompt engineering. At run-time, we assemble prompts dynam-
 364 ically based on the user’s question and context, which ensures
 365 the generation of relevant and factual responses. This is similar
 366 to recent approaches such as retrieval-augmented generation [24],
 367 where the model receives additional knowledge or documents with
 368 the query and has to extract the relevant fragments.

369 We first provide an overall description of the setup, this is the
 370 same for all prompts. We then follow with an in-depth description
 371 of the particular user location (i.e., “*The user is in a room with three*
 372 *doors, one of them pink to the right...*”) and the relevant upcoming
 373 steps. Next, we append the original instruction given to the user
 374 along with commands to restrain the response (i.e., “*This is the*
 375 *instruction given to the user, do not repeat it back*”) and the history
 376 of what the user has done and said. Finally, we provide several
 377 examples of how to respond to user questions briefly, similar to
 378 in-context learning [4, 27].

379 Throughout the prompt, we condition the model to generate
 380 short responses that keep to the factual description of the user’s
 381 location and actions. We rely on the LLM to retrieve the relevant
 382 knowledge in the context (i.e., room description) and use it to
 383 guide the user in the environment. Providing this additional context
 384 is crucial to reduce the LLM’s tendency to hallucinate facts in
 385 responses [28]. This is however an open issue with LLMs and a
 386 limitation of our system, which could sometimes produce irrelevant
 387 responses.

3.3 Virtual Environment

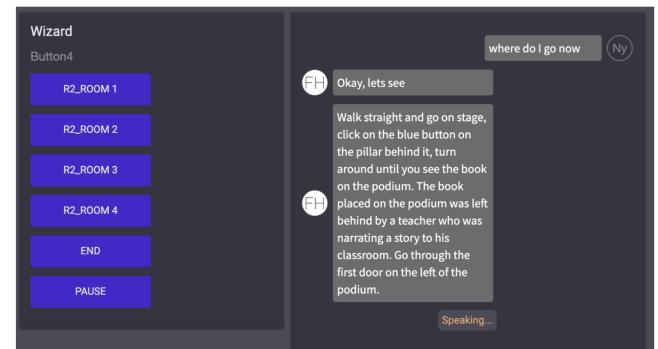
390 We used the video game Minecraft as our virtual environment for
 391 the system. This popular video game runs on most laptop devices,
 392 the controls are intuitive and its code is easy to modify. Its main
 393 feature is that everything is made of blocks (walls, plants, etc.) so
 394 it is ideal to create visually-rich scenarios with controlled conditions.
 395 Figure 5 shows how the environment is made of these square
 396 ‘blocks’.

397 Within Minecraft, we built intricate houses with distinctive
 398 rooms and corridors to navigate between them. Each room was de-
 399 signed with particular features, puzzles and decorations that serve
 400 both as distractions and landmarks (see Figure 5 for an indoor gar-
 401 den). We carefully balanced the ambiguity of the rooms versus their

403 ⁶<https://platform.openai.com/docs/models/gpt-3-5-turbo>, last accessed: 20/05/2024

404 ⁷<https://platform.openai.com/docs/models/gpt-3-5-turbo>, last accessed: 20/05/2024

405 ⁸usually 0.75 seconds



406 **Figure 4: The Furhat wizarding interface with buttons on the**
 407 **left to localise the LLM’s knowledge to the specific room. We**
 408 **also keep track of the conversation on the right side.**

423 difficulty (i.e., rooms may have several flowers of different colours,
 424 or doors located on opposite sides). Unique decorations are also key
 425 to creating narratives around them, although we avoid items that
 426 are too rare and give strong impressions at first sight (i.e., large,
 427 colourful objects). We instead opt for objects commonly found in
 428 houses or Minecraft, such as bookshelves, tables or paintings.

429 Each room, except the final one, has three exit doors scattered
 430 leading to exit corridors. Only one of these doors would be the
 431 ‘correct’ door to follow according to the instructions, but users can
 432 go through any of them. The corridors are all identical and long
 433 so that the user can only see another door at the end and has to
 434 commit to crossing one of these to continue the task. Regardless
 435 of the exit corridor that the user chooses to cross, we programmed
 436 the game to teleport the user’s avatar to the same correct corridor
 437 when mid-way through. In other words, users are unaware of any
 438 wrong decisions they made whilst following the instructions in the
 439 environment, and we can continue the experiment normally. In the
 440 end, users will traverse the same four rooms regardless of which
 441 paths they choose.



442 **Figure 5: An example of a room with lush greenery in the**
 443 **virtual environment, Minecraft. One of the exit corridors is**
 444 **on the left with the door open.**

Narrative	Non-Narrative	
In the first room, walk besides the meeting table to the wall with the pink flowers. These pink flowers are tulips, I picked them from a garden and planted them because I love how they smell. Press all the buttons above the pink flowers, then walk until you see a pink door and go through it.	In the first room, walk to the wall with the pink flowers, press all the buttons above the pink flowers, then walk until you see a pink door and go through it.	523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
In the second room, walk straight and go on the podium, click on the blue button on the pillar behind it, turn around until you see the book on the podium. The book placed on the podium was left behind by a teacher who was narrating a story to his classroom. Go through the first door on the left of the podium.	In the second room, walk straight until you see the podium, click on the blue button on the pillar behind it, turn around until you see the book on the podium then go through the first door on the left.	541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580

Figure 6: Sample instruction triplets (*Room, Puzzle, Door*) for the Narrative and Non-Narrative conditions.

We built two different houses, with unique navigation routes, in the virtual environment to balance any effects related to the within-subject format of our study (§4.2). While one of the key features of Minecraft is its building mechanics, once the environment had been built by the researchers we changed the game settings to avoid participants modifying the built environment.

3.4 Interaction

The interaction with the system begins with the robot verbally greeting users, and outlining the task to complete. There is a chance to have some chit-chat (“*How are you?*”) and, after confirming that they understand the overall task, Furhat proceeds to read out loud the full pre-written instruction set for the task. The Furhat repeats the instructions as many times as needed, after which users can start navigating the virtual environment. Up to this point, the conversation uses Furhat’s NLU and DM, plus our handcrafted intents and rules.

During the environment navigation, users can engage in free-form voice interaction with Furhat at any point. This allows for spontaneous and informal questions about the task, instructions or directions. Our system is always aware of the user’s position, so prompts are dynamically assembled to retrieve the most relevant description as in, the current room or previous actions. We could not establish a direct connection between the virtual environment and Furhat, so one of the facilitators acts as the Wizard of Oz and manually presses buttons to track room changes or actions in the wizarding interface (see Figure 4).

Throughout the interaction, Furhat moves its head every so often as to keep a natural interaction and follows the user with its gaze. When it is speaking, Furhat also makes facial gestures related to what is being said.

4 EXPERIMENTAL DESIGN

The main aim of this pilot experiment was to assess the system’s functionality and learn what works in the interaction process. As

previously mentioned (§2), human-human conversations can benefit from using personal narratives in instruction-following tasks. We see improvements in recall and engagement [36], although to our knowledge few works have evaluated this in human-robot scenarios. Therefore, we design into our pilot study a comparison of narratives in a navigation task using the system in Section 3. Users would receive an instruction from the embodied agent, with or without narratives added to it, and then would have to navigate the virtual environment based on these instructions. We hypothesise that including narratives would improve the recall of the route instructions and the likeability of the embodied agent. We also anticipate that the enhanced recall will lead to fewer queries during the open navigation.

This section first provides details about the instructions and narratives used, followed by the study’s methodology, a short description of the data and participants and finishes with a brief discussion of the results.

4.1 Narratives vs Non-Narratives Instructions

We evaluate two conditions in which we modify the route instructions: **Narrative** and **Non-Narrative**, see Figure 6 for sample instructions.

Each route instruction consists of a set of triplets (*Room, Puzzle, Door*) with details about the current room being discussed, the actions the user should do to complete a short puzzle in the room, and then a reference to the exit door, respectively. These detail chains have a mix of referring expressions with objects, visual properties, spatial references and actions [34]. We heavily relied on colours as they provided a balance between uniqueness and ambiguity (e.g., there could be several objects of the same colour, but only one is the ‘blue flower’) and avoided instructions that were too specific. Our initial tests suggested that simply providing references to the exit door was trivial to recall, thus we included short puzzles that users have to complete before continuing (e.g., “*press the yellow buttons*”) in each room as distractions.

Finally, we concatenate a triplet for each room, up to four, for the final route instructions, leading to potentially lengthy overall instructions. We do not expect users to be able to recall the complete instruction set, hence they have to engage in short dialogues with Furhat as the guide or oracle during the navigation. Users are encouraged to request information or ask follow-up questions without restrictions, which may range from clarificational exchanges to refresh what they could remember (e.g., “Was it the door next to the yellow or red flowers?”) to more direct ones (e.g., “Where do I go now?”). The system has a record of the user’s location and previous actions so it can tailor responses to their context. A sample dialogue can be found in Table 1.

For the **Narrative** instructions condition, we add short personal stories to one of the referring expressions akin to those used in human-human conversations. For example, after mentioning ‘pink flowers’, we add ‘These are tulips, I picked them from a garden...’. We manually crafted these narratives based on previous works that use a mix of human experiences (memories, actions, smells...).

Table 1: Sample dialogue collected between the system and a participant during the navigation.

Participant Enters Room 1	
USR	I am in the first room what should I do now?
SYS	Pull the lever down above the beige door on the green wall. It should emit a red light. Go to the beige door on the purple wall, located between the bookshelf and the blue flowers. Go through that door.
USR	You mean the door is right to the TNT?
SYS	Yes, the door with the lever is to the right of the TNTs
Participant Enters Room 2	
USR	Can you describe the second room ?
SYS	The second room has green, yellow, blue, and orange walls. Press all yellow buttons on the blue wall. Exit next to the American Gothic painting.
USR	Where is the wall?
SYS	The blue wall is opposite the orange wall and has buttons in the middle.
USR	Which painting?
SYS	The American Gothic painting, located on the right side of the yellow wall.
	...

4.2 Methodology

We assess the system and explore the effects of narratives in a guided navigation task, for which we have two conditions: **Narrative** and **Non-Narrative** route instructions.

We set up the Furhat and virtual environment in a controlled lab (see Figure 3). The table had the Furhat on the right and a monitor showing the virtual environment in full screen along with a keyboard and mouse in the centre. There were also two other laptops on the table for the facilitators to monitor and control the study, but participants only interacted with the keyboard and mouse and could not see the content of the other two laptops during the interaction period.

We first gave a brief to participants about the experimental task, followed by a consent form and a short pre-study survey with demographic questions, which also checked their experience with the video game Minecraft before. We then taught participants Minecraft’s basic controls and how it works in a pre-study room. Once they were familiar with the setup, environment and Furhat, they could start the experiment.

We ran a within-subjects study, so participants tried both conditions. To avoid familiarity biases with the route on the second

condition, we used two different routes in the virtual environment so that instructions and navigation would be unique. We counterbalanced the study by randomly assigning participants to a route (R1 or R2) and either the **Narrative** or **Non-Narrative** conditions. Thus, participants always navigated both routes, but the combination of Condition×Route would be distinct (each with unique instruction sets), controlling for potential changes in perceived map difficulty and maintaining two study conditions.

Upon starting the task, participants talked with Furhat, who provided the full route instructions. Participants could ask Furhat to repeat them if they wished as they could be lengthy. Participants then navigated through the rooms, asking questions if needed, until they reached the final room. As previously mentioned (§3.3), participants were not aware of wrong turns or actions (puzzles were not mandatory to continue), so they always reached the end regardless of the route.

After each navigation, participants completed a post-condition survey with questions regarding:

- **Recall:** a free-text open question that measures recall [40]. We divide the answers into context elements, for which each element scores separately. We also use gist-based recall[40], accepting synonyms (i.e., close wording scores positively), but only counting correct elements once. Due to the open-ended nature of this question, participants often included additional details not in the original instructions. We excluded these details in the analysis and ignored the order of context elements.
- **Perceived Usefulness:** a series of questions adapted from the ALMERE model [17] on a 7-point Likert scale.
- **Perceived Ease of Use:** a series of questions adapted from the ALMERE model [17] on a 7-point Likert scale.
- **Likeability:** a subscale of the Godspeed questionnaire [2] on a 5-point scale of opposing concepts.

The full experiment lasts around 20 minutes, with 5 to 8 minutes of navigation in each route.

4.2.1 Measurements. We also collected several objective metrics during the participant’s navigation, the number of: **incorrect door** (goes through incorrect exit door); **puzzle errors** (incorrectly or incomplete room puzzle); and **navigation questions** (requests information about where to go next).

During the interaction, participants occasionally asked about Minecraft or how to do something in the environment. Although Furhat was able to answer these correctly (possibly due to GPT-3.5’s large knowledge base and training data), we excluded these from any analysis in this paper.

4.3 Participants

We recruited 24 participants (7 female, 17 male, most commonly in the age range of 18–22 years) through internal mailing lists. Regarding the familiarity with the virtual environment, 58% of the participants had played Minecraft in the past, 25% knew of it but had never played it, and 17% had never heard of it before. This study received approval from our institution’s Ethics Board.

Table 2: Descriptive statistics for all numeric measures ($N=23$). Cronbach's α is used to measure internal consistency of the sub-scale items, where a higher value denotes higher measured internal consistency (\uparrow).

		Median	Mean	Std. Deviation	Minimum	Maximum	Cronbach's α (\uparrow)
701 702 703 704 705 706 707 708 709 710 711 712 713 714 715	IncorrectDoor	NonNarrative Narrative	1.000 0.000	0.696 0.435	0.703 0.728	0.000 0.000	2.000 2.000
	PuzzleErrors	NonNarrative Narrative	0.000 0.000	0.696 0.913	0.876 1.125	0.000 0.000	3.000 3.000
701 702 703 704 705 706 707 708 709 710 711 712 713 714 715	FollowUpQuestions	NonNarrative Narrative	3.000 2.000	2.565 2.435	1.121 1.199	0.000 0.000	5.000 5.000
	LikeabilityScore	NonNarrative Narrative	4.400 4.800	4.304 4.443	0.638 0.671	3.200 3.000	5.000 5.000
701 702 703 704 705 706 707 708 709 710 711 712 713 714 715	PercievedUsefulnessScore	NonNarrative Narrative	6.000 6.000	5.812 6.043	0.904 0.727	3.667 4.333	7.000 7.000
	PercievedEaseOfUseScore	NonNarrative Narrative	6.200 6.200	6.130 6.078	0.759 0.822	4.600 4.400	7.000 7.000
701 702 703 704 705 706 707 708 709 710 711 712 713 714 715	RecallNavigation	NonNarrative Narrative	0.176 0.400	0.247 0.341	0.218 0.253	0.000 0.000	0.667 0.700

4.4 Results

One outlier was removed from the analysis due to metrics being more than 1.5 box-lengths from the edge of the boxplot box. Cronbach's Alpha was calculated over the Godspeed Questionnaire subscales, shown in Table 2. Internal consistency of all measured subscales is > 0.7 , which is deemed acceptable.

Trends were visible in the data, with the narrative condition indicating a generally lower mean of incorrect doors taken and follow up questions, alongside a higher rate of recall, likeability, and perceived usefulness ratings. A lower indicated mean ease of use score, and higher number of puzzle errors was however found for the narrative condition. Due to the limitations that will be covered in Section 6, no result was statistically significant by way of a repeated-measures MANOVA.

5 CONTRIBUTIONS

The main contributions of this paper are summarised as follows:

- (1) We use narratives with an embodied robot agent for a navigation task, combining Human-Robot Interaction and conversational AI in an ecologically valid task. Previous studies have not explored the use of narratives in these embodied settings as far as we are aware.
- (2) We carry out an initial pilot study with in-person participants and assess the feasibility of running larger-scale experiments. We evaluate how humans engage with the agent providing personal stories, finding some indication that participants prefer narratives and that they improve recall.
- (3) We provide a set of guidelines so future works can use this as a starting point when designing studies with virtual environments in social robotics. Environments focused on HRI are often time-consuming to adapt to a particular task if they are not constrained or outdated. We instead propose using easily-modifiable video games as test beds for future studies as they are faster to adapt at the cost of fidelity.

6 DISCUSSION

In this section, we first examine the metric results from our pilot study, and comment on how we will change this for a larger evaluation. We continue by covering the limitations of this study, which when coupled with lessons learned from the study, allow us to construct design guidelines for future studies aiming to use simple videogame-based virtual environments for navigation studies.

6.1 Summary of results

Trends indicate that narratives may be preferred in terms of recall, usefulness, likeability, and in terms of reducing the number of incorrect doors taken and follow-up questions asked. However, trends also indicate that the number of puzzle errors may increase, and the perceived ease of use score may increase. These trends do align with our existing hypotheses, and from this, we have created a list of research questions to investigate further in a larger study:

- **RQ1:** How does including narratives affect the recall of route instructions?
- **RQ2:** Does the use of narratives influence the perceived likeability, usefulness, and ease of use of the robot?
- **RQ3:** What impact does the use of narratives have on the occurrence and type of follow-up questions?

We hypothesise that while narratives will increase perceived usefulness scores, it will decrease perceived ease of use score due to the increased length of text provided by the narrative condition. In addition, we hypothesise that there will be a correlation between number of follow-up questions, and incorrect turns taken, based on an increased recall rate from the narrative instruction set.

6.2 Limitations

Due to the insignificant results, it is clear there are limitations present in the study design. Of most immediate importance is the sample size, which being $N=24$ is simply too low. A power analysis performed in G*Power[12] estimates that a total sample size of > 70 may be required to see an effect on certain variables. This is based on

813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
an estimated effect size based on the mean and standard deviation of differences. A higher effect size may be possible by addressing further issues, such as the homogeneity of the age of participants, or by adjusting the links between narratives and landmarks - however more work is required.

6.3 Design Guidelines for Simple Virtual Environments

822
823
824
825
826
Throughout the pilot and testing phase of our developed system, we found a number of design choices that either had been made or must be made in future virtual environment studies for them to be successful. We encapsulate these guidelines and useful ideas in the following points:

- (1) Unlike physical environments, the nature of our virtual environment and teleportation corridors meant that we could always control the rooms that participants entered. This problem of control is present in other studies looking at navigation in virtual environments. [29] compares learning conditions in a virtual environment, and finds no group differences in a route retracing recall task. They did however have the issue that when a participant took a wrong turn or missed a turn, the experimenter had to stop them, which we hope can be solved in future using these more modern virtual environments capable of teleporting participants discretely.
- (2) Striking the right balance between objects too abstract or ambiguous and objects that were too specific to the game environment was difficult. Considering that 17% of our participants had never heard of Minecraft before, and 25% had heard of it but never played it, objects that were very obvious to those who had played the game could not be used, for example, a “trapdoor.” Similarly, trying to create objects that did not exist in the game, but could be created with other objects, often were too ambiguous for users to recognise. A balance had to be struck, using objects that were apparent and present both in the game and the real world, such as bookshelves or paintings.
- (3) Localising the LLM to avoid unneeded context helped reduce hallucinations greatly. This reduced noise meant that the LLM could properly have conversations about what mattered at the time, and would not confuse rooms with one another.
- (4) Prompting the LLM with very specific instructions, and detailing every object in the room - coupled with the point above - reduced the number of hallucinations to practically zero.

7 CONCLUSION AND FUTURE WORK

In this paper, we create a system capable of giving directions to navigate a Minecraft virtual environment, using an LLM to answer any follow-up questions the user may ask. We used prompt engineering to localise the LLM to specific rooms along the route, meaning specific questions referencing the environment could be answered effectively. We then added narratives to the route instructions in an attempt to explore their impact on social robot acceptance factors and recall with a human study. We evaluate the feasibility of using

871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
an embodied agent as the instruction giver for the virtual environment in an initial pilot experiment and find that participants tend to prefer an agent that uses personal stories. Further work will explore the impact of these factors with a larger pool of participants and fix shortcomings of the current design. Finally, we provide design guidelines for future works in HAI with virtual environments based on lessons learnt during the development of the system. Further research is needed to uncover the role of narratives in navigational tasks with embodied social robots, particularly as they become more integrated into everyday scenarios.

REFERENCES

- [1] Gary L. Allen. 1997. From knowledge to words to wayfinding: Issues in the production and comprehension of route directions. In *Spatial Information Theory A Theoretical Basis for GIS (Lecture Notes in Computer Science)*, Stephen C. Hirtle and Andrew U. Frank (Eds.). Springer, Berlin, Heidelberg, 363–372. https://doi.org/10.1007/3-540-63623-4_61
- [2] Christoph Bartneck, Dana Kulic, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 71–81. <https://doi.org/10.1007/s12369-008-0001-3>
- [3] Hilary L. Bekker, Anna E. Winterbottom, Phyllis Butow, Amanda J. Dillard, Deb Feldman-Stewart, Floyd J. Fowler, Maria L. Jibaja-Weiss, Victoria A. Shaffner, and Robert J. Volk. 2013. Do personal stories make patient decision aids more effective? A critical review of theory and evidence. *BMC Medical Informatics and Decision Making* 13, 2 (Nov. 2013), S9. <https://doi.org/10.1186/1472-6947-13-S2-S9>
- [4] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models Are Few-Shot Learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada) (NIPS’20). Curran Associates Inc., Red Hook, NY, USA, Article 159, 25 pages.
- [5] Jerome Bruner. 1986. *Actual Minds, Possible Worlds*. Harvard University Press. Google-Books-ID: iDANAQAMAAJ.
- [6] David Caduff and Sabine Timpf. 2005. The Landmark Spider: Representing Landmark Knowledge for Wayfinding Tasks. 30–35.
- [7] Neeraj Cherakara, Finny Varghese, Sheena Shabana, Nivan Nelson, Abhiram Karukayil, Rohith Kulothungan, Mohammed Afil Farhan, Birthe Nesset, Meriam Moujahid, Tanvi Dinkar, Verena Rieser, and Oliver Lemon. 2023. FurChat: An Embodied Conversational Agent using LLMs, Combining Open and Closed-Domain Dialogue with Facial Expressions. In *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Svetlana Stoyanov, Shafiq Joty, David Schlangen, Ondrej Dusek, Casey Kennington, and Malihi Alikhani (Eds.). Association for Computational Linguistics, Prague, Czechia, 588–592. <https://doi.org/10.18653/v1/2023.sigdial-1.55>
- [8] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeonjae Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thamanulayan Sankaranarayana Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. 2022. PaLM: Scaling Language Modeling with Pathways. arXiv:2204.02311 [cs.CL]
- [9] Helen Couclelis. 1996. Verbal Directions for Way-Finding: Space, Cognition, and Language. In *The Construction of Cognitive Maps*, Juval Portugali (Ed.). Springer Netherlands, Dordrecht, 133–153. https://doi.org/10.1007/978-0-585-33485-1_7
- [10] Maartje MA de Graaf, Somaya Ben Allouch, and Jan AGM van Dijk. 2018. A phased framework for long-term user acceptance of interactive technology in domestic environments. *New Media & Society* 20, 7 (July 2018), 2582–2603. <https://doi.org/10.1177/1461444817727264> Publisher: SAGE Publications.

- [11] Michel Denis, Francesca Pazzaglia, Cesare Cornoldi, and Laura Bertolo. 1999. Spatial discourse and navigation: an analysis of route directions in the city of Venice. *Applied Cognitive Psychology* 13, 2 (1999), 145–174. [https://doi.org/10.1002/\(SICI\)1099-0720\(199904\)13:2<145::AID-ACP550>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1099-0720(199904)13:2<145::AID-ACP550>3.0.CO;2-4) _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291099-0720%28199904%2913%3A2%3C145%3A%AAID-ACP550%3E3.0.CO%3B2-4>.
- [12] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39, 2 (May 2007), 175–191. <https://doi.org/10.3758/BF03193146>
- [13] Gary L. Allen. 1999. Spatial abilities, cognitive maps, and wayfinding : Bases for individual differences in spatial cognition and behavior. *Wayfinding Behavior* (1999), 46–80. <https://cir.nii.ac.jp/crid/1573105975561296512> Publisher: The Johns Hopkins University Press.
- [14] Arthur C. Graesser, Kathy Hauft-Smith, Andrew D. Cohen, and Leonard D. Pyles. 1980. Advanced Outlines, Familiarity, and Text Genre on Retention of Prose. *The Journal of Experimental Education* 48, 4 (1980), 281–290. <https://www.jstor.org/stable/20151355> Publisher: Taylor & Francis, Ltd.
- [15] Arthur C. Graesser, Nicholas L. Hoffman, and Leslie F. Clark. 1980. Structural components of reading time. *Journal of Verbal Learning & Verbal Behavior* 19 (1980), 135–151. [https://doi.org/10.1016/S0022-5371\(80\)90132-2](https://doi.org/10.1016/S0022-5371(80)90132-2) Place: Netherlands Publisher: Elsevier Science.
- [16] Melanie C. Green, Jeffrey J. Strange, and Timothy C. Brock (Eds.). 2002. *Narrative impact: Social and cognitive foundations*. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US. Pages: xv, 373.
- [17] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2010. Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. *International Journal of Social Robotics* 2, 4 (Dec. 2010), 361–375. <https://doi.org/10.1007/s12369-010-0068-5>
- [18] Paivi Heikkila, Hanna Lammi, Marketta Niemela, Kathleen Belhassen, Guillaume Sarthou, Antti Tammela, Aurelie Clodic, and Rachid Alami. 2019. Should a Robot Guide Like a Human? A Qualitative Four-Phase Study of a Shopping Mall Robot. In *Social Robotics (Lecture Notes in Computer Science)*, Miguel A. Salichs, Shuzhi Sam Ge, Emilia Ivanova Barakova, John-John Cabibihan, Alan R. Wagner, Alvaro Castro-Gonzalez, and Hongsheng He (Eds.). Springer International Publishing, Cham, 548–557. https://doi.org/10.1007/978-3-030-35888-4_51
- [19] Amanda Hinnant, María E. Len-Rios, and Rachel Young. 2013. JOURNALISTIC USE OF EXEMPLARS TO HUMANIZE HEALTH NEWS. *Journalism Studies* 14, 4 (2013). <https://doi.org/10.1080/1461670X.2012.721633>
- [20] Höök, Kristina. 1991. *An Approach to a Route Guidance Interface*. Licentiate Thesis. Stockholm University, Department of Computer and Systems Sciences.
- [21] Yunfan Jiang, Agrim Gupta, Zichen Zhang, Guanzhi Wang, Yongqiang Dou, Yanjun Chen, Li Fei-Fei, Anima Anandkumar, Yuke Zhu, and Linxi Fan. 2023. VIMA: General Robot Manipulation with Multimodal Prompts. In *Fortieth International Conference on Machine Learning*.
- [22] Maria L. Jibaja-Weiss, Robert J. Volk, Thomas S. Granchi, Nancy E. Neff, Emily K. Robinson, Stephen J. Spann, Noriaki Aoki, Lois C. Friedman, and J. Robert Beck. 2011. Entertainment education for breast cancer surgery decisions: A randomized trial among patients with low health literacy. *Patient Education and Counseling* 84, 1 (July 2011), 41–48. <https://doi.org/10.1016/j.pec.2010.06.009>
- [23] Austin S. Kilaru, Jeanmarie Perrone, Catherine L. Auriemma, Frances S. Shofer, Frances K. Barg, and Zachary F. Meisel. 2014. Evidence-based Narratives to Improve Recall of Opioid Prescribing Guidelines: A Randomized Experiment. *Academic Emergency Medicine* 21, 3 (2014), 244–249. <https://doi.org/10.1111/acem.12326> _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/acem.12326>
- [24] Patrick Lewis, Ethan Perez, Aleksandr Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 9459–9474. https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780c1bc26945df7481e5-Paper.pdf
- [25] Mei Yili Lim, David A. Robb, Bruce W. Wilson, and Helen Hastie. 2023. Feeding the Coffee Habit: A Longitudinal Study of a Robo-Barista. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 1983–1990. <https://doi.org/10.1109/RO-MAN57019.2023.10309621> ISSN: 1944-9437.
- [26] Daniel Little. 2010. *New Contributions to the Philosophy of History*. Springer Science & Business Media. Google-Books-ID: ZTmV7Kas3aUC.
- [27] Pengfei Liu, Weizhe Yuan, Jinlai Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-Train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Comput. Surv.* 55, 9, Article 195 (jan 2023), 35 pages. <https://doi.org/10.1145/3560815>
- [28] Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On Faithfulness and Factuality in Abstractive Summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 1906–1919. <https://doi.org/10.18653/v1/2020>
- [29] Chiara Meneghetti and Francesca Pazzaglia. 2021. Navigating in Virtual Environments: Does a Map or a Map-Based Description Presented Beforehand Help? *Brain Sciences* 11, 6 (June 2021), 773. <https://doi.org/10.3390/brainsci11060773>
- [30] Pierre-Emmanuel Michon and Michel Denis. 2001. When and Why Are Visual Landmarks Used in Giving Directions?. In *Spatial Information Theory (Lecture Notes in Computer Science)*, Daniel R. Montello (Ed.). Springer, Berlin, Heidelberg, 292–305. https://doi.org/10.1007/3-540-45424-1_20
- [31] Michelle Miller-Day and Michael L. Hecht. 2013. Narrative Means to Preventative Ends: A Narrative Engagement Framework for Designing Prevention Interventions. *Health communication* 28, 7 (Oct. 2013), 657–670. <https://doi.org/10.1080/10410236.2012.762861>
- [32] Sewon Min, Xinxia Lyu, Ari Holtzman, Mikel Artetxe, Mike Lewis, Hannaneh Hajishirzi, and Luke Zettlemoyer. 2022. Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 11048–11064. <https://doi.org/10.18653/v1/2022.emnlp-main.759>
- [33] H. Morris. 1993. Narrative representation, narrative enactment, and the psychoanalytic construction of history. *International Journal of Psycho-Analysis* 74, 1 (1993), 33–54.
- [34] Per Persson. 1998. Supporting Navigation in Digital Environments: A Narrative Approach. In *Exploring Navigation: Towards a Framework for Design and Evaluation of Navigation in Electronic Spaces*. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=94c3864c68f946cffc5881b9dd4139122536f70a>
- [35] Martin Raubal and Stephan Winter. 2002. Enriching Wayfinding Instructions with Local Landmarks. In *Geographic Information Science (Lecture Notes in Computer Science)*, Max J. Egenhofer and David M. Mark (Eds.). Springer, Berlin, Heidelberg, 243–259. https://doi.org/10.1007/3-540-45799-2_17
- [36] Victoria A. Shaffer, Elizabeth S. Focella, Andrew Hathaway, Laura D. Scherer, and Brian J. Zikmund-Fisher. 2018. On the Usefulness of Narratives: An Interdisciplinary Review and Theoretical Model. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine* 52, 5 (April 2018), 429–442. <https://doi.org/10.1093/abm/kax008>
- [37] Victoria A. Shaffer and Brian J. Zikmund-Fisher. 2013. All stories are not alike: a purpose-, content-, and valence-based taxonomy of patient narratives in decision aids. *Medical Decision Making: An International Journal of the Society for Medical Decision Making* 33, 1 (Jan. 2013), 4–13. <https://doi.org/10.1177/0272989X12463266>
- [38] Robert J. Volk, Maria L. Jibaja-Weiss, Sarah T. Hawley, Suzanne Kneuper, Stephen J. Spann, Brian J. Miles, and David J. Hyman. 2008. Entertainment education for prostate cancer screening: a randomized trial among primary care patients with low health literacy. *Patient Education and Counseling* 73, 3 (Dec. 2008), 482–489. <https://doi.org/10.1016/j.pec.2008.07.033>
- [39] Jason Wei, Maarten Bosma, Vincent Y. Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V. Le. 2022. Finetuned Language Models Are Zero-Shot Learners. [arXiv:2109.01652 \[cs.CL\]](https://arxiv.org/abs/2109.01652)
- [40] Michael B. W. Wolfe and Joseph A. Mienko. 2007. Learning and memory of factual content from narrative and expository text. *The British Journal of Educational Psychology* 77, Pt 3 (Sept. 2007), 541–564. <https://doi.org/10.1348/000709906X143902>

Received 7 June 2024

acl-main.173
987
Chiara Meneghetti and Francesca Pazzaglia. 2021. Navigating in Virtual Environments: Does a Map or a Map-Based Description Presented Beforehand Help? *Brain Sciences* 11, 6 (June 2021), 773. <https://doi.org/10.3390/brainsci11060773>
988
989
Pierre-Emmanuel Michon and Michel Denis. 2001. When and Why Are Visual Landmarks Used in Giving Directions?. In *Spatial Information Theory (Lecture Notes in Computer Science)*, Daniel R. Montello (Ed.). Springer, Berlin, Heidelberg, 292–305. https://doi.org/10.1007/3-540-45424-1_20
990
991
992
Michelle Miller-Day and Michael L. Hecht. 2013. Narrative Means to Preventative Ends: A Narrative Engagement Framework for Designing Prevention Interventions. *Health communication* 28, 7 (Oct. 2013), 657–670. <https://doi.org/10.1080/10410236.2012.762861>
993
994
995
Sewon Min, Xinxia Lyu, Ari Holtzman, Mikel Artetxe, Mike Lewis, Hannaneh Hajishirzi, and Luke Zettlemoyer. 2022. Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 11048–11064. <https://doi.org/10.18653/v1/2022.emnlp-main.759>
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044