

Experts' Perception of AI-Generated Personalised Stories for Health Promotion

Jayalakshmi BASKAR^{a,1}, Vera C. KAELIN^a and Helena LINDGREN^a

^aDepartment of Computing Science, Umeå University, 901 87 Umeå, Sweden

ORCID ID: Jayalakshmi Baskar <https://orcid.org/0000-0001-7267-6534>, Vera C.

Kaelin <https://orcid.org/0000-0003-1290-9441>, Helena Lindgren

<https://orcid.org/0000-0002-8430-4241>

Abstract. A qualitative study is presented aimed to explore how person-tailored stories generated by a storytelling system agent built on a combination of argumentation theory and large language models (LLM) can support a person in reflecting on activities promoting health and potential lifestyle changes. We conducted a formative evaluation with ten domain experts to explore perceived purposes, benefits and risks. They saw both potential opportunities, if applied with a clear purpose for their target groups, and risks in using LLM-based stories for health reflection if not carefully designed to adapt to an individual's needs. The results will guide further design and implementation of the storytelling functionality, and studies involving older adults.

Keywords. Storytelling, Personalisation, Artificial Intelligence, Argumentation Theory, Behaviour change, Health promotion, Co-design, Large language models

1. Introduction

Encouraging reflection on daily activities and their effects on health and wellbeing is an important aspect of promoting lifestyle changes that support healthy living [1]. Technologies designed to foster such reflection can enhance autonomy, social participation, and the establishment of healthy routines [2]. However, many existing digital interventions rely mainly on reminders or factual feedback [3], and therefore risk missing the emotional and narrative dimensions that make reflection meaningful [4].

Storytelling is a fundamental way through which humans communicate, reflect upon and construct their understanding of reality [5]. In medicine, narrative has also been recognized as a valuable instrument for communicating with patients [6]. Within Artificial Intelligence (AI), research increasingly focuses on the narrative and social aspects of interacting with agentic systems [7]. Large language models (LLMs) offer new opportunities to generate adaptive and personalised stories which, when grounded in a person's habits, daily activities, needs, and goals, may serve as a tool for reflecting on how lifestyle changes could improve health and prevent disease. Nevertheless, the recent development of LLMs challenges the traditional expectations on believability. Although these models can generate fluent stories, their probabilistic nature may lead to outputs that are readable but not factually or logically grounded - commonly referred to

¹ Corresponding Author: Jayalakshmi Baskar, jaya@cs.umu.se

as *hallucinations* [8]. It is worth noting that human storytelling similarly incorporates elements of creativity to make stories more engaging and personally meaningful.

When combined with AI technologies that ground the purposes of stories in a formal framework of knowledge, LLM-generated stories may achieve a balance between perceived accuracy and creative flexibility. In this study, we apply *argumentation theory*, which provides a structured approach to sense-making by framing the interaction as an argumentation-based dialogue with a clear purpose, such as reasoning about and drawing conclusions from past activities (inquiry), planning activities (deliberation), or challenge existing habits to promote healthier behaviour (persuasion) [9].

The aim of this study was to explore the potential risks, benefits, and purposes of an interactive storytelling system that integrates LLMs with computational argumentation to promote autonomy, health, and wellbeing across different populations, with a particular focus on older adults. Specifically, our main research question was: How do experts in occupational therapy, physiotherapy, and nursing perceive the usefulness, risks, creativity, and the potential of AI-generated person-tailored stories to motivate and promote reflection on daily activities to improve health?

This paper presents the results of a formative evaluation study involving domain experts. We examined how different *dialogue types* (inquiry, deliberation, and persuasion) and *creativity* (hallucination) levels (low and high) are perceived with respect to *purpose*, e.g., to support lifestyle changes to improve health and prevent illnesses, and how these perceptions vary when imagining different user groups (e.g., older adults and younger people with cognitive difficulties).

2. Method

Ten domain experts (eight occupational therapists, one physiotherapist, and one nurse) from three European countries participated in a qualitative evaluation study, part of a participatory design process. All participants had substantial experience designing or evaluating digital support for daily life and health-related activities among people with cognitive or physical challenges.

A prototype of a personalised storytelling system was developed as a web application in Python, embedding a storytelling module (hereafter, the storytelling agent) [10]. The agent collects and stores information about the user's activities. The storytelling agent retrieves the user's logged activities and generates a user model. The agent then generates eight stories in English that connect a person's logged activities with four pre-defined prompts designed to target four generic purposes related to promoting reflection on meaningful activities for improving health and well-being (Table 1). For story generation, we used OpenAI ChatGPT-5-mini [11] accessed via API. Each story was produced in two creativity modes (low, high) and three dialogue types representing different communicative intents (inquiry, deliberation, and persuasion).

The study was conducted as follows. Prior to participation, each expert received detailed information about the study and provided informed consent. Sessions were held online or in person, based on participant preference. As a first step, each participant identified their own activities within four predefined categories: physical activities, social activities, recovery activities, and everyday activities. They then assessed these activities in terms of their perceived importance, fun, the motives for doing them, and the frequency with which they had engaged in each activity over the past weeks.

Table 1. Prompts, dialogue types, and intended purposes used for story generation.

Purpose	Type	Prompt assumed to be posted by a user
Assessment of past activities	Inquiry	“Did I do something important the past week?”
Assessing fulfilment of needs and goals	Inquiry	“Tell me how I have fulfilled my motives the past week/month.”
Decide what to do and plan the coming week	Deliberation	“What should I do to make the coming week more recovering than the past week?”
Change attitude to do something differently	Persuasion	“Please convince me what to do next week so it becomes a happy week.”

In the next step, the system generated eight different stories based on their activities and the prompts. The story variants (high and low creativity) were presented in random order. Semi-structured interviews were conducted during their use of the system. Given the exploratory purpose of the study, the data analysis was conducted qualitatively.

3. Results

The experts approached the task from their own perspective (e.g., busy, middle-aged with children), or from the perspective of someone in their envisioned user group (older adults, young people with neuropsychiatric diagnoses or developmental disorders), and they switched between the perspectives during the session.

All experts viewed the storytelling function as a promising tool for supporting reflection and engagement among their clients. For example, they envisioned how AI-generated stories could provide cognitive support by helping clients **plan the upcoming week** and **maintain daily structure** (e.g., organizing tasks and remembering what to do). Additionally, one participant mentioned that such stories could support individuals discover new activities, for instance, to reduce isolation or increase physical activity.

However, their expectations and preferred applications differed. The perceived usefulness of a story depended on the right balance between the type of dialogue and its purpose. Inquiry stories (focused on reflecting on what had been important during the past week) were seen to potentially be useful for memory support among older adults, but less so for younger adults, who were believed to focus more on planning the future. This story type was considered relevant mainly when a person explicitly sought to reflect on their motives. Deliberation stories (focused on deciding what to do in the coming week) were judged most helpful when they offered small, concrete suggestions that supported planning. Persuasion stories (focused on encouraging certain behaviour or attitudes to promote well-being) were perceived as useful when they were concise and grounded in the person's own activities; when the phrasing became figurative, experts reported confusion. For example, experts appreciated persuasion stories when they offered grounded and concrete suggestions, such as *“Combine tasks when it helps, like a walk with the kids that doubles as fresh air and company, or a quick grocery stop before cooking together.”* Such phrasing was described as realistic and easy to relate to, in contrast to more figurative language like *“Imagine grocery aisles as quiet treasure hunts—choosing bright vegetables, a loaf of bread that smells like comfort—each selection a tiny promise you keep to yourself.”* Several experts found them poetic but harder to follow or connect to everyday experience, especially for older adults or individuals with cognitive challenges.

Comments about perceived inconsistencies in the stories were mainly about (i) contextual drift, words or examples that were not relevant to the person's logs, (ii)

summary distortions, over generic statements about the week, and (iii) stylistic phrasing, long, figurative sentences. They raised bother and reduced clarity. For example, participants noted that when stories described their motives unprompted, it sometimes created a sense of resistance, for instance, by making them present the activity as more fun than it really was. Similarly, one participant noted, when a user rates an activity as highly important, the story should reflect that importance rather than express uncertainty. Experts preferred brief, literal wording and where appropriate, concrete pointers to the underlying activities from the logs to make the stories' claims feel grounded. The experts also requested clear, simple and concrete language with short sentences and minimal stylish flourish, especially for individuals with cognitive difficulties or dementia symptoms.

The preferences were also linked to some risks of AI-generated stories that the participating experts mentioned, such as potential loss of comprehension due to figurative wording (long sentences), and issues of grounding and trust when statements lacked clear connection to a person's activity logs.

4. Discussion

This study explored the purposes, potential risks, and benefits of an interactive storytelling system that integrates LLMs with computational argumentation to promote autonomy, health, and well-being. As part of a participatory design approach, interviews with experts revealed that they saw promise in such a system for supporting reflection and engagement among their clients. The experts described how stories employing different dialogue types (i.e., inquiry, deliberation, and persuasion) could serve distinct purposes, each with its own strengths and limitations. These findings suggest that it is important to clarify the individual's goal for using the storytelling system so that the chosen dialogue type aligns with those goals. For example, if the goal is memory support, inquiry-based stories may be most suitable, whereas deliberation-focused stories may be more meaningful for offering suggestions and motivating future activities.

The narrative styles and levels of creativity were experienced differently across experts. While most preferred low-creativity stories with literal wording, some favoured the more creative versions. However, the results also highlighted that high-creative stories may risk becoming detached from the person's lived experiences. The heterogeneity in preferences for story types may stem from differences in personal expectations, cultural backgrounds, and individuals' perceived connections between story content and their everyday activities. These results suggest that personalisation should encompass not only the content of the stories but also their style. Similarly, prior research has shown that older adults exhibit different narrative styles when describing their activity performance to an agent [12]: some provide precise step-by-step accounts of their actions, while others include elaborations and additional context - an approach that may parallel the high-creativity stories observed in this study.

A major limitation of the study was the low number of participating domain experts as well as the storytelling system not being presented in participants' native language. However, the group of participants provided diverse viewpoints and constructive insights in how AI-generated storytelling may be used and experienced by different groups and individuals, fulfilling the study's formative and exploratory purpose. As a next step, a group of older adults will be involved in further forming and evaluating the storytelling functionality, which will be further developed based on the results in this study.

5. Conclusions

This paper presents a formative participatory study that explored how domain experts perceived AI-generated personalised stories designed to promote reflection in daily activities. The findings highlight that purpose alignment is important for their perceived usefulness. Low-creativity stories with clear, literal phrasing were preferred for older adults and persons with cognitive difficulties, while high-creativity narratives were seen as engaging but occasionally confusing. The participating experts emphasized that story type should fit the individual's goal: deliberative stories for planning and organisation, inquiry-based stories for memory support, and persuasive stories only when grounded in the person's actual activities. Across dialogue types, clarity, concise and exact use of words and contextual grounding emerged as key to credibility and motivation.

To summarise, both potential opportunities were envisioned, if applied with a clear purpose for their target groups, and risks in using AI-based stories for health reflection if not carefully designed to adapt to an individual's needs. These insights underline the value of combining argumentation structures with LLMs, and to apply participatory design to make person-tailored storytelling both reliable and personally meaningful.

References

- [1] Lim CY, Berry AB, Hartzler AL, Hirsch T, Carrell DS, Bermet ZA, Ralston JD. Facilitating self-reflection about values and self-care among individuals with chronic conditions. In: Brewster S, Fitzpatrick G, editors. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*; 2019 May 4-9; Glasgow, Scotland UK. New York (NY): Association for Computing Machinery. p. 1-12.
- [2] Lindgren H, Kilic K. Participatory design and evaluation of digital coaching for improving health—the star multicomponent lifestyle intervention. *Front Digit Health*. 2025 Jun; 7:1-21.
- [3] De Witte NA, Joris S, Van Assche E, Van Daele T. Technological and digital interventions for mental health and wellbeing: an overview of systematic reviews. *Front Digit Health*. 2021 Dec; 3:1-10.
- [4] Green MC, Brock TC. The role of transportation in the persuasiveness of public narratives. *J Pers Soc Psychol*. 2000 Nov;79(5):701-21.
- [5] Bruner J. The narrative construction of reality. *Crit Inq*. 1991 Oct;18(1):1-21.
- [6] Palla I, Turchetti G, Polvani S. Narrative Medicine: theory, clinical practice and education—a scoping review. *BMC Health Serv Res*. 2024 Sep;24(1):1-18.
- [7] Dautenhahn K. The art of designing socially intelligent agents: Science, fiction, and the human in the loop. *Applied artificial intelligence*. 1998 Oct;12(7-8):573-617.
- [8] Aher GV, Arriaga RI, Kalai AT. Using large language models to simulate multiple humans and replicate human subject studies. In: Krause A, Brunskill E, Cho K, Engelhardt, Sabato S, Scarlett J. *Proceeding of the 40th International Conference on Machine Learning*; 2023 Jul 23-29; Honolulu, Hawaii. San Diego, CA: International Conference on Machine Learning. p. 337-371.
- [9] Walton D, Krabbe EC. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. Albany (NY): State University of New York Press; 1995.
- [10] Baskar J, Kilic K, Kaelin VC, Lindgren H. Towards collaborative planning for health promotion through person-tailored storytelling and argumentation. In: Braccini M, De Filippo A, Milano M, Saffiotti A, Vallati M. *CEUR Workshop Proceedings of the 1st Workshop on Human-AI Collaborative Systems co-located with 28th European Conference on Artificial Intelligence (ECAI 2025)*; 2025 Oct 25-30; Bologna, Italy. p. 70-83.
- [11] OpenAI. Introducing gpt-5, August 7 2025. <https://openai.com/index/introducing-gpt-5/>
- [12] Lindgren H, Kaelin VC, Ljusback AM, Tewari M, Persiani M, Nilsson I. To adapt or not to adapt? older adults enacting agency in dialogues with an unknowledgeable agent. In: Boratto L, Gena C, Marras, editors. *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*; 2024 July 1-4; Cagliari, Italy. New York, NY: Association for Computing Machinery. p. 307-316.