Using Generative AI to Produce Situated Action Recommendations in Augmented Reality for High-Level Goals

BRENNAN JONES, Meta Reality Labs Research, Redmond, WA, USA, brennanj@meta.com
YAN XU, Meta Reality Labs Research, Redmond, WA, USA, yanx@meta.com
MARY ANNE HOOD, Meta Reality Labs Research, Redmond, WA, USA, maryanne@meta.com
MOHAMMAD SHAHIDUL KADER, Meta Reality Labs Research, Redmond, WA, USA, mohammadk1@meta.com
HAMID EGHBALZADEH, Meta Reality Labs Research, Redmond, WA, USA, heghbalz@meta.com

Many people pursue long-term **high-level goals** related to areas such as fitness, mental health, or skill enhancement. Additionally, each individual pursuing a high-level goal may pursue it differently than others, depending on their **contexts**. In this research, we explore the use of generative AI for producing **context-dependent and situated recommendations in augmented reality (AR)** for actions that individuals can take toward their high-level goals. We developed a technology probe and ran a user study in a mock home environment to understand how users experience and perceive such recommendations in the context of a home environment and the tools contained within it. We found that users **value the passive nature of such recommendations**, and the **roles that they can play to inform, motivate, and inspire the user**. However, compared to receiving advice from friends, family, and experts, users are still skeptical about AI-generated suggestions, particularly for critical and sensitive goals, and there are still design considerations to explore to understand how to improve the content and delivery of such recommendations.

1 INTRODUCTION

People pursuing **high-level goals** (i.e., complex long-term goals such as *lose weight* or *learn a new language*) that are novel to them may not know where to begin their goal journey. Additionally, each individual pursuing a high-level goal may pursue it differently than others, depending on their **contexts** (e.g., time, place, available resources).

In this work, we explore the use of generative AI models such as large language models (LLMs) to **recommend actions that individuals can take toward their high-level goals** and **situate those recommendations within the right contexts** (e.g., at the relevant time/place or next to the relevant tools). We envision a future where people wear augmented-reality (AR) glasses that are feasible to wear all-day everyday and that are at least as ubiquitous in people's lives as smartphones are today. Such AR devices can **capture information about the user and their context**, and **use that information to recommend relevant actions to the user and their choice of goals**.

2 BACKGROUND

Large Language Models. LLMs, such as GPT-3 [2], OPT [32], and PaLM [9], are natural language models pre-trained with large amounts of text that generate and predict human-like text based on a prompt or series of prompts given to it. Many of such models are task-agnostic, and have been applied to activities such as summarizing text [21], generating code [3, 6–8, 28], programming robots [4], and health consultation [29]. While these tasks involve helping a user accomplish a low-level goal (i.e., a short-term goal or immediate task), our work is focused on how such models could recommend actions for high-level long-term goals.

Recommender Systems. Recommender systems strive to provide users with suggestions that are most relevant to the user at a given moment or context [26]. Traditional recommendation techniques include *collaborative filtering* [25], (recommending by matching a user with other users) and *content-based filtering* [20] (recommending based on the user's previous activities). Researchers have also explored context-based techniques such as recommending suggestions

Fig. 1. The AR prototype displaying *situated action recommendations* next to relevant objects in a mock apartment space.

based on the user's mood [5]. Recommendation systems have been applied to domains such as online shopping [19] and recipe discovery [16, 24]. Recent work (e.g., [33]) has begun to explore the use of generative AI models such as LLMs for generating recommendations. However, there has been little research on recommender systems that produce suggestions for behavior change and goal pursuit.

Technologies for Behavior Change and Goal Pursuit. Previous research has explored *persuasive technologies* [15], or technologies designed to invoke behavior change, habit formation, and help the user pursue high-level goals (e.g., [10, 11, 15]). Research on such tools has suggested that they should be designed to "consider the **practical constraints** of users' lifestyles" [10], which can include individuals' **contexts** (such as their daily schedules, home environments, and the resources in their homes) and **affordances** (such as their personal skills and capabilities given their contexts). In this research, we are exploring how generative AI models can take users' contexts and affordances into consideration to recommend relevant and timely actions toward users' high-level goals.

Context. Context is not always easy to define, neither is its role in the design and experience of ubiquitous computing systems [13]. However, a user's context plays a role in defining how they can act toward their goals [14]. For example, if a user is driving, they cannot follow a recommendation to *"do five minutes of squats"*. Therefore, a user's context can determine how they react to recommendations (e.g., [12, 17, 22, 23]). In this research, we recognize that context can include a variety of factors, including (but not limited to): *available resources* (e.g., tools in the home), *time*, *location*, *state of one's environment*, *current task*, one's *goals*, *personal attributes*, and one's *physiological* or *psychological state*.

3 TECHNOLOGY PROBE AND USER STUDY IN A MOCK APARTMENT

As an initial step in this research, we focus primarily on the context of *available resources*, more specifically on the tools available in one's environment. We designed a technology probe [18] (Figure 1) that explores the use of situated action

recommendations in AR to **help users discover action possibilities** for pursuing their high-level goals within the contexts of the objects contained in their living spaces. This system displays **situated action recommendations** for the user's goal pursuit, triggered whenever certain objects are available to the user (e.g., within the user's view).

This prototype was implemented in Unity [27] and runs on the Microsoft HoloLens 2. At present, it tracks objects in a mock apartment space by detecting fiducial markers using the Vuforia Engine [1]. The recommendations that the prototype presents are displayed in the scene near the relevant objects. Each recommendation has (1) an **action name** (e.g., *"use the yoga mat to do yoga"*) and a list of the user's **goals** that the action supports (e.g., *"improve fitness"*, *"improve mental health"*). In the current prototype, the recommendations were pre-generated using GPT-3 and manually added to a JSON configuration database used by the prototype.

We ran a user study in a mock studio apartment, where we began to address the following research questions:

(1) How do people **experience AI-generated context-dependent and situated action recommendations**?
   (a) How do these experiences **compare to individuals' current approaches** to seeking and receiving advice and motivation for their goal pursuits?
(2) How do users **perceive advice generated by AI compared to advice from other sources** (e.g., close social ties, domain experts)?
(3) How can AI–generated situated action recommendations in AR **be delivered better**, such that users are more likely to find them useful for information, motivation, or inspiration?

We ran a total of 39 participants, including 25 females and 14 males, aged 19 to 73 (M = 38, SD = 14). All participants experienced the recommendations in the study as coming from generative AI. However, for *Research Question 2*, we sought to understand how participants would experience the recommendations if they perceived them as coming from *friends/family* or *experts*. Thus, we implemented a between-subjects design where for 14 participants, the AI-generated recommendations were labeled as authored by the participant's *friends/family* (e.g., tagged with one of their friend's or family member's names), and for 13 other participants they were labeled as authored by *experts* (e.g., tagged with the name and credentials of a fictional domain expert, e.g., *"Dr. Chloe Brown, Licensed Mental Health Counselor"*). Even though all participants knew the recommendations were AI-generated, they were asked to imagine as though they were 'authored' by these people. For the remaining 12 participants, the recommendations were not labeled with an author name, and participants purely perceived the recommendations as generated by AI. The composition of participants across the three groups was balanced so that each group had roughly the same distribution in terms of *how much they tended to trust smart recommendations*, their *previous experience using AR*, and their *ages* and *genders*.

At the beginning of the study session, participants filled out a brief survey, where among other things, they selected **three goals** from a list of seven (which included *Improve fitness*, *Be more eco-friendly*, *Tidy up the home*, *Improve mental health*, *Connect with friends*, *Learn a new language*, and *Learn a new skill*) that they were either actively pursuing or genuinely cared about. Participants were then interviewed briefly about their selected goals. Following this, each participant completed three trials of an activity where they used the prototype to walk around the apartment, view recommendations for their goals, and **choose three recommendations to accept**. For each trial, the prototype displayed recommendations for a different combination of **two of their three selected goals**. Following the activity, participants were interviewed a final time, where they discussed their experiences using the prototype.

Full details about the study procedure and findings will be published in a future paper.

3.1 Preliminary Findings

Our study findings produced the following insights:

INSIGHT 1: Users value that situated action recommendations are passive (delivered with little effort), thus providing high convenience and saving time. Participants valued that the recommendations came passively, and that they would not need to actively seek out the advice as they normally would need to do when asking an expert or a friend for advice. Some participants compared these passive recommendations to existing environmental cues that they use as reminders to do an activity — e.g., one participant mentioned using dirty dishes in her kitchen sink as an 'environmental cue' to wash the dishes. Situated recommendations could serve to nudge the user to perform activities that do not normally have such natural environmental cues. Participants mentioned that this could save them time, or be useful in situations where they do not have a lot of mental energy to brainstorm ideas for actions to take.

INSIGHT 2: While users are skeptical of AI-generated suggestions, they value their potential to improve action discoverability. Most participants mentioned that they trust experts and their close social ties more for critical domains like mental health and fitness. Moreover, participants felt that advice from their friends and family was more personalized to them than AI-generated advice. However, participants valued AI's potential to help them discover a larger and more creative set of action ideas. Some participants mentioned that they would consider narrowing down or filtering out AI-generated suggestions based on the opinions of their close social ties or experts.

INSIGHT 3: Users find suggestions most useful for unfamiliar actions and goals, but are more likely to adopt familiar actions. Participants tended to accept actions that they were familiar with (e.g., actions that they typically already do) or that they anticipated did not require high effort or friction. As has been seen in previous research on song selection choices in music streaming [30], while participants found unfamiliar choices to be interesting and valued the diversity of the choices being offered, they ultimately gravitated toward existing habits and choosing familiar options. Reasons for doing this that participants mentioned include that **familiar actions involve low effort and time commitment**, and that the user **already knows the action works for them**. There could be design choices that might encourage users to select less familiar but beneficial options more frequently, including providing information to the user about the *usefulness* and *'friction level'* of an action, to increase its familiarity to the user.

4 CONCLUSION AND FUTURE RESEARCH DIRECTIONS

Our work has begun to explore how generative AI, with an understanding of the user's goals and contexts, can **produce action recommendations for the user's goals** and **situate them at the relevant time and place in AR**, thus delivering suggestions that are more grounded, specific, and timely than traditional approaches. While our preliminary insights reveal that participants value this style of advice, future work should explore how to improve the user experience of AI-generated situated action recommendations.

For instance, our insights reveal that suggestions can be delivered with more details about the expected effort, time commitment, and how to perform the action. We plan to further explore **what other information and delivery styles allow users to benefit more from AI-generated suggestions**.

We also plan to explore **how users can interact with the underlying model itself** in order to tailor it to produce better outputs, thus **giving the user more agency and partnership with the AI model**. Given that it has been shown that generative AI can produce better outputs through back-and-forth reasoning and conversation with the user [31], it would be worthwhile to explore how such mixed-initiative interactions can help the user improve the content and delivery of the recommendations output by the model, to make them more relevant and timely to the user.

REFERENCES

[1] [n. d.]. Vuforia Enterprise Augmented Reality (AR) Software | PTC. https://www.ptc.com/en/products/vuforia

[2] 2021. GPT-3 Powers the Next Generation of Apps. https://openai.com/blog/gpt-3-apps/

[3] 2021. OpenAI Codex. https://openai.com/blog/openai-codex/

[4] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Daniel Ho, Jasmine Hsu, Julian Ibarz, Brian Ichter, Alex Irpan, Eric Jang, Rosario Jauregui Ruano, Kyle Jeffrey, Sally Jesmonth, Nikhil J. Joshi, Ryan Julian, Dmitry Kalashnikov, Yuheng Kuang, Kuang-Huei Lee, Sergey Levine, Yao Lu, Linda Luu, Carolina Parada, Peter Pastor, Jornell Quiambao, Kanishka Rao, Jarek Rettinghouse, Diego Reyes, Pierre Sermanet, Nicolas Sievers, Clayton Tan, Alexander Toshev, Vincent Vanhoucke, Fei Xia, Ted Xiao, Peng Xu, Sichun Xu, Mengyuan Yan, and Andy Zeng. 2022. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. https://doi.org/10.48550/arXiv.2204.01691 arXiv:2204.01691 [cs].

[5] Ivana Andjelkovic, Denis Parra, and John O'Donovan. 2016. Moodplay: Interactive Mood-based Music Discovery and Recommendation. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization (UMAP '16)*. Association for Computing Machinery, New York, NY, USA, 275–279. https://doi.org/10.1145/2930238.2930280

[6] Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, and Charles Sutton. 2021. Program Synthesis with Large Language Models. https://doi.org/10.48550/arXiv.2108.07732 arXiv:2108.07732 [cs].

[7] Sid Black, Leo Gao, Phil Wang, Connor Leahy, and Stella Biderman. 2022. GPT-Neo: Large Scale Autoregressive Language Modeling with Mesh-Tensorflow, 2021. *URL: https://doi. org/10.5281/zenodo* 5297715 (2022).

[8] Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, Dave Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William Hebgen Guss, Alex Nichol, Alex Paino, Nikolas Tezak, Jie Tang, Igor Babuschkin, Suchir Balaji, Shantanu Jain, William Saunders, Christopher Hesse, Andrew N. Carr, Jan Leike, Josh Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba. 2021. Evaluating Large Language Models Trained on Code. https://doi.org/10.48550/arXiv.2107.03374 arXiv:2107.03374 [cs].

[9] 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, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan 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. https://doi.org/10.48550/arXiv.2204.02311 arXiv:2204.02311 [cs].

[10] Sunny Consolvo, Katherine Everitt, Ian Smith, and James A. Landay. 2006. Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. Association for Computing Machinery, New York, NY, USA, 457–466. https://doi.org/10.1145/1124772.1124840

[11] Sunny Consolvo, David W. McDonald, and James A. Landay. 2009. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, New York, NY, USA, 405–414. https://doi.org/10.1145/1518701.1518766

[12] Xiang Ding, Jing Xu, Honghao Wang, Guanling Chen, Herpreet Thind, and Yuan Zhang. 2016. WalkMore: promoting walking with just-in-time context-aware prompts. In *2016 IEEE Wireless Health (WH)*. 1–8. https://doi.org/10.1109/WH.2016.7764558

[13] Paul Dourish. 2004. What we talk about when we talk about context. *Personal and Ubiquitous Computing* 8, 1 (Feb. 2004), 19–30. https://doi.org/10.1007/s00779-003-0253-8 Number: 1.

[14] BJ Fogg. 2009. A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology (Persuasive '09)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/1541948.1541999

[15] B. J. Fogg. 2002. Persuasive technology: using computers to change what we think and do. *Ubiquity* 2002, December (Dec. 2002), 5:2. https://doi.org/10.1145/764008.763957

[16] Kazjon Grace, Elanor Finch, Natalia Gulbransen-Diaz, and Hamish Henderson. 2022. Q-Chef: The impact of surprise-eliciting systems on food-related decision-making. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3491102.3501862

[17] Maxi Heitmayer and Saadi Lahlou. 2021. Why are smartphones disruptive? An empirical study of smartphone use in real-life contexts. *Computers in Human Behavior* 116 (March 2021), 106637. https://doi.org/10.1016/j.chb.2020.106637

[18] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. 2003. Technology probes: inspiring design for and with families. In

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03). Association for Computing Machinery, New York, NY, USA, 17–24. https://doi.org/10.1145/642611.642616

[19] Brandon Huynh, Adam Ibrahim, Yun Suk Chang, Tobias Höllerer, and John O'Donovan. 2018. A Study of Situated Product Recommendations in Augmented Reality. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 35–43. https://doi.org/10.1109/AIVR.2018.00013

[20] Hosein Jafarkarimi, Alex Tze Hiang Sim, and Robab Saadatdoost. 2012. A naive recommendation model for large databases. *International Journal of Information and Education Technology* 2, 3 (2012), 216. Publisher: IACSIT Press.

[21] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. https://doi.org/10.48550/arXiv.1910.13461 arXiv:1910.13461 [cs, stat].

[22] Chung Tse Liu, Steen J. Hsu, and Chia Tai Chan. 2014. Context-Aware Prompting System for Improving Physical Activity. In *2014 International Symposium on Computer, Consumer and Control*. 836–839. https://doi.org/10.1109/IS3C.2014.221

[23] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My Phone and Me: Understanding People's Receptivity to Mobile Notifications. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1021–1032. https://doi.org/10.1145/2858036.2858566

[24] Azzurra Pini, Jer Hayes, Connor Upton, and Medb Corcoran. 2019. AI Inspired Recipes: Designing Computationally Creative Food Combos. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312948

[25] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2011. Introduction to Recommender Systems Handbook. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor (Eds.). Springer US, Boston, MA, 1–35. https://doi.org/10.1007/978-0-387-85820-3_1

[26] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2022. Recommender Systems: Techniques, Applications, and Challenges. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, New York, NY, 1–35. https://doi.org/10.1007/978-1-0716-2197-4_1

[27] Unity Technologies. [n. d.]. Unity Real-Time Development Platform | 3D, 2D VR & AR Engine. https://unity.com/

[28] Priyan Vaithilingam, Tianyi Zhang, and Elena L. Glassman. 2022. Expectation vs. Experience: Evaluating the Usability of Code Generation Tools Powered by Large Language Models. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3491101.3519665

[29] Lu Wang, Munif Ishad Mujib, Jake Williams, George Demiris, and Jina Huh-Yoo. 2021. An Evaluation of Generative Pre-Training Model-based Therapy Chatbot for Caregivers. https://doi.org/10.48550/arXiv.2107.13115 arXiv:2107.13115 [cs].

[30] Morgan K. Ward, Joseph K. Goodman, and Julie R. Irwin. 2014. The same old song: The power of familiarity in music choice. *Marketing Letters* 25, 1 (March 2014), 1–11. https://doi.org/10.1007/s11002-013-9238-1

[31] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. https://doi.org/10.48550/arXiv.2201.11903 arXiv:2201.11903 [cs].

[32] Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, Todor Mihaylov, Myle Ott, Sam Shleifer, Kurt Shuster, Daniel Simig, Punit Singh Koura, Anjali Sridhar, Tianlu Wang, and Luke Zettlemoyer. 2022. OPT: Open Pre-trained Transformer Language Models. https://doi.org/10.48550/arXiv.2205.01068 arXiv:2205.01068 [cs].

[33] Yuhui Zhang, Hao Ding, Zeren Shui, Yifei Ma, James Zou, Anoop Deoras, and Hao Wang. 2021. Language models as recommender systems: Evaluations and limitations. In *NeurIPS 2021 Workshop on I (Still) Can't Believe It's Not Better*. https://www.amazon.science/publications/language-models-as-recommender-systems-evaluations-and-limitations