A Cautionary Tale About AI-Generated Goal Suggestions

Falk Lieder
Max Planck Institute for Intelligent Systems, Tübingen
Germany
falk.lieder@tuebingen.mpg.de

Pin-Zhen Chen
Max Planck Institute for Intelligent Systems, Tübingen
Germany
pin-zhen.chen@tuebingen.mpg.de

Jugoslav Stojcheski
Max Planck Institute for Intelligent Systems, Tübingen
Germany
jugoslav.stojcheski@gmail.com

Saksham Consul
Max Planck Institute for Intelligent Systems, Tübingen
Germany
sakshamconsul@gmail.com

Viktoria Pammer-Schindler
Graz University of Technology
Austria
viktoria.pammer-schindler@tugraz.at

ABSTRACT

Setting the right goals and prioritizing them might be the most crucial and the most challenging type of decisions people make for themselves, their teams, and their organizations. In this article, we explore whether it might be possible to leverage artificial intelligence (AI) to help people set better goals and which potential problems might arise from such applications. We devised the first prototype of an AI-powered digital goal-setting assistant and a rigorous empirical paradigm for assessing the quality of AI-generated goal suggestions. Our empirical paradigm compares the AI-generated goal suggestions against randomly-generated goal suggestions and unassisted goal-setting on a battery of self-report measures of important goal characteristics, motivation, and usability in a large-scale repeated-measures online experiment. The results of an online experiment with 259 participants revealed that our intuitively compelling goal suggestion algorithm was actively harmful to the quality of the people's goals and their motivation to pursue them. These surprising findings highlight three crucial problems to be tackled by future work on leveraging AI to help people set better goals: i) aligning the objective function of the AI algorithms with the design goals, ii) helping people quantify how valuable different goals are to them, and iii) preserving the user's sense of autonomy.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Interactive systems and tools**; **Laboratory experiments**; **Usability testing**.

KEYWORDS

AI alignment, productivity tools, goal-setting, prioritization

ACM Reference Format:
Falk Lieder, Pin-Zhen Chen, Jugoslav Stojcheski, Saksham Consul, and Viktoria Pammer-Schindler. 2022. A Cautionary Tale About AI-Generated Goal Suggestions. In *Mensch und Computer 2022 (MuC '22), September 4–7, 2022, Darmstadt, Germany*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3543758.3547539

MuC '22, September 4–7, 2022, Darmstadt, Germany

ACM ISBN 978-1-4503-9690-5/22/09.
https://doi.org/10.1145/3543758.3547539

1 INTRODUCTION

Goals are an essential driver of productivity and progress. Each person typically juggles about 15 goals at any given time [16]. Given that people have to juggle so many goals in parallel, it comes as no surprise that they often struggle to identify and prioritize what is most important for their goals and projects in the long run. Instead, they often get side-tracked by less important tasks that are more urgent in the short run. This can derail critical long-term projects, which is frustrating for individuals, their managers, and their organizations. While existing productivity tools help people remember their tasks and goals, they rarely help people prioritize their most important goals.

The few to-do list apps (e.g., Amazing Marvin) and goal-setting apps (e.g., Complice) that do support prioritization put the burden of figuring out which goals are most important on the user. This is also true of research on supporting goal-setting with chatbots [e.g., 2, 7, 10]. The few systems that leverage artificial intelligence (AI) to help the user make better decisions [1, 3, 9, 19, 24] are only applicable in specific domains, such as emergency management and business decisions. Recent work has begun to fill this gap by leveraging planning algorithms to generate prioritized to-do lists that incentivize each task according to how valuable it is for the user's long-term goals [5, 14, 15, 25]. Experiments show that these methods can help people overcome motivational obstacles to working on a goal they have already selected [14]. However, there is no empirical work on whether and, if so, how they can be used to help people discover and select new goals that are even better than those they are currently pursuing. More generally, there is virtually no work on leveraging the superhuman performance of machine planning [21, 23] to support human goal-setting. We, thus, know almost nothing about the potential risks and benefits of these potential future applications of AI.

Our long-term goal is to create intelligent digital companions that help people set and achieve socially-valuable goals. Here, we explore how AI might be used to support goal-setting and what can go wrong. To test this approach, we developed an early prototype of an intelligent digital goal-setting assistant and a rigorous empirical methodology for assessing the quality of AI-generated goal suggestions and the usefulness and usability of our digital goal-setting assistant. The plan for this paper is as follows. We first describe the first prototype of our AI-powered digital goal-setting assistant, then present the results of an experiment we conducted to evaluate it. We close by reflecting on the risks and challenges of AI-generated goal suggestions and highlighting important directions for future research.

2 A FIRST ATTEMPT AT LEVERAGING AI TO COMPUTE GOAL SUGGESTIONS

To explore whether AI can be used to safely and reliably improve human goal-setting, we adapted the intelligent planning algorithms [5, 25] developed to help people achieve their current goals into a method for helping them decide which goals to pursue next. Here, we used those algorithms to create a minimum viable version of an intelligent digital goal-setting assistant. Our design goal was to create a web application that gives the user a suggestion for which of their goals they might want to prioritize next month. If the AI algorithms achieve their purpose, the suggested goals should, on average, be more beneficial to the user than the goals they would have chosen without the digital assistant.

The first prototype of our digital goal-setting assistant comprises three main components: i) a graphical user interface for setting goals and milestones (see Figure 1a), ii) AI-powered algorithms for computing goal suggestions, and iii) the goal-suggestion page (see Figure 1b). Our digital assistant first asks the user about the most crucial thing they want to get done by the end of their life and a rough time estimate of how much effort it would take to achieve that goal. The user's answer becomes their first candidate goal. Its value is set to 100% and serves as a reference point for the values of other candidate goals. On the subsequent goal-setting page, shown in Figure 1a, the user can break down their first goal into a series of milestones. Each milestone should be a major step toward achieving the goal. The user can also add additional goals. For each additional goal, the interface asks the user to estimate its value relative to the crucial life goal they shared in the first step (e.g., 10% as valuable). Moreover, our application asks the user to provide a rough estimate of the completion time of each goal and milestones in hours. The user can set as many goals and milestones as they like. When the user is done entering their goals and milestones, they click on the button labeled "Which of these goals should I focus on first?". Our web application then adds an "everything else" task to each goal and sends the user's anonymous data to a server that uses the AI algorithms developed by Stojcheski et al. [25] to compute optimal goal suggestions. The time estimate of the "everything else" task is the difference between the user's time estimate for the goal and the sum of the user's time estimates for the goal's milestones. While the server computes the goal suggestions, the website says, "Please wait while our AI computes its suggestions." This usually takes about 2-3 seconds. Once the point values computed by the to-do list gamification server have been received, our web application generates the goal-suggestions page (see Figure 1b). Our web application ranks the milestones by how many points they are worth per hour of work required to achieve them. It then recommends the user to first prioritize the goal corresponding to the highest-ranked milestone and to then work on the goal corresponding to the second highest-ranked milestone (see Figure 1b). The ranked list of all other goals is shown under the heading "Your other goals are". As illustrated in Figure 1b), the goal-suggestions page also shows each goal's milestones, values, and time estimates.

3 EXPERIMENT: HOW GOOD ARE THE SUGGESTED GOALS?

Our goal in developing the goal-suggestion application was to help people select weekly goals that are useful for achieving their long-term objectives. We further assumed that such goals would be more interesting than the presumably more mundane short-term goals that people might prioritize otherwise. Therefore, the primary goal of our experiment was to test whether AI-generated goal suggestions can help people choose goals that are useful (H1), interesting (H2), and far-sighted (H3). Our second goal was to assess the user experience with the first prototype of our goal-setting assistant and identify improvement opportunities. We predicted that users would rate our prototype to be more useful than a placebo version that selects its goal suggestions at random (H4a). Moreover, we predicted that users of our digital goal-setting assistant would report more positive and fewer negative emotions (H4b) and be more willing to use it again (H4c) than users who received random goal suggestions.

3.1 Methods

We ran an experiment that randomly assigned participants to one of three goal-setting interventions that provided AI-powered goal suggestions (experimental condition), randomly selected goal suggestions (active control condition), and no goal suggestions (passive control condition), respectively. We compared these three goal-setting interventions in terms of how much they improved the quality of the user's goals.

Participants. We recruited 86 participants per condition on the online study platform Prolific (210 female, 43 male; age: 18–50+). Two basic attention-check questions tested if participants read the questions and understood the information. We excluded 7.3% of all participants who failed one or both attention-checks. Participants in the experimental or active control group were paid £3 for about 30 minutes of work. Participants in the passive control condition were paid £1.5 for about 15 minutes of work. Participants who passed all attention-checks received an additional performance-dependent bonus of £1.2 (experimental condition and active control condition) or £0.6 (passive control condition).

Materials. We instructed participants to use our web application to set goals for the next month. The instructions explained the necessary steps and the information participants they were asked to provide (i.e., goals, milestones, goal values, and time estimates). To proceed to the web application, participants had to pass a quiz about those instructions.

We measured user experience with the meCue questionnaire [18]. This includes measures of our application's perceived usefulness (H4a), user's emotional experience (H4b), and their intention to use our application again in the future (H4c). To measure what kinds of goals our AI-powered digital assistant suggested, we asked

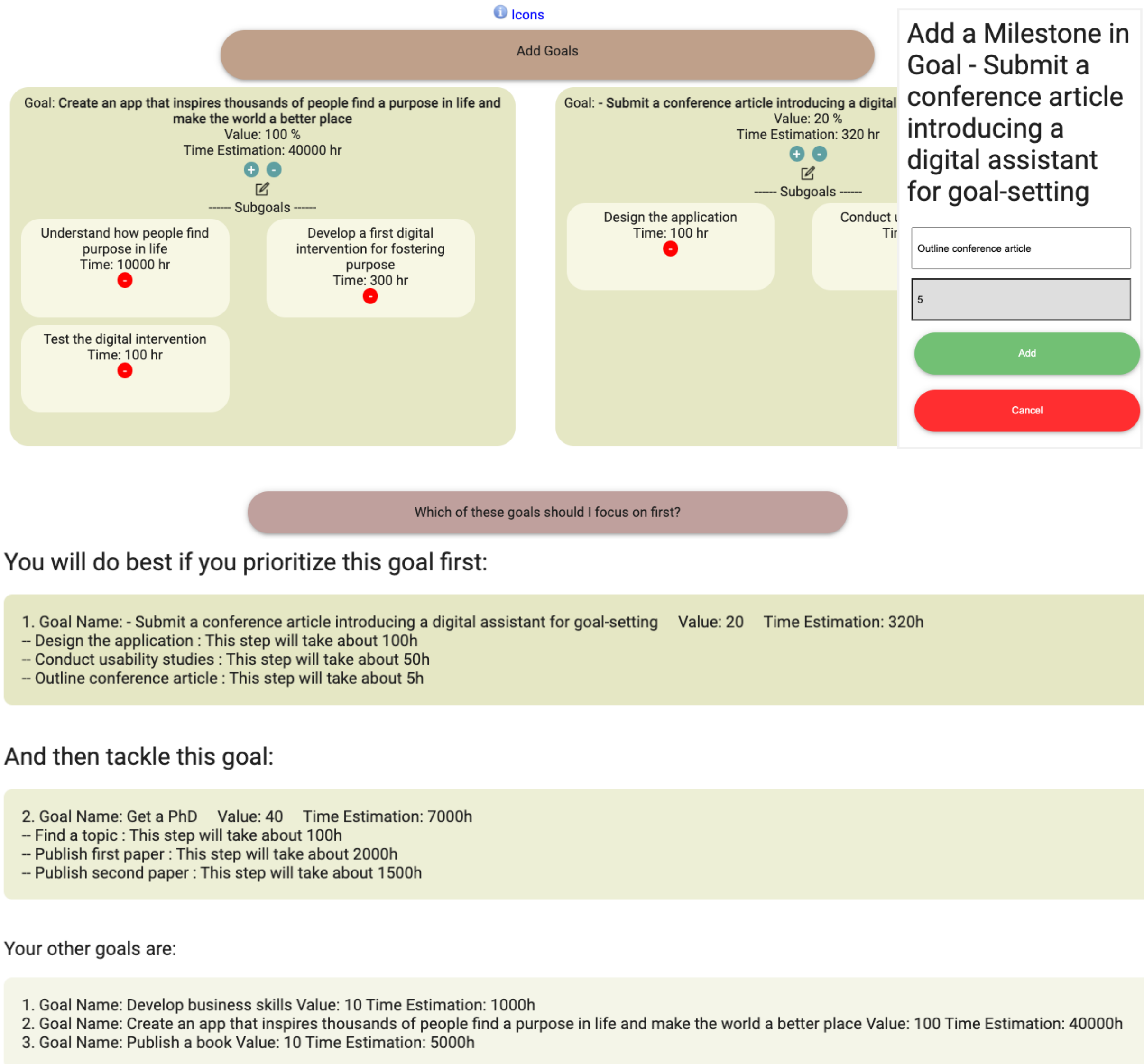

Figure 1: Screenshots of the goal-setting interface (a) and the goal suggestion page (b) of our digital goal-setting assistant.

participants to rate them on the selected subscales of the Goal Characteristics Questionnaire (GCQ) [11] summarized in Table 1. To measure the user's satisfaction with those suggestions, we asked them to rate their quality and the value of potential improvements on two 9-point Likert scales with 8 and 5 items, respectively.

For goal suggestions to be effective, the users have to be motivated to pursue the suggested goals. Therefore, we devised a three-item Likert scale, self-report questionnaire measuring how likely users are to pursue the suggested goal, how motivated they are to do so, and how strongly they want to prioritize it. In addition, we measured the quality of the user's motivation to pursue the goal using the self-concordance scale [22]. Moreover, we asked them to estimate how many hours they were planning to invest in the goal next week. Finally, to investigate whether our application might be more beneficial or more appealing for certain types of users, we measured participants' propensity to plan [17].

Table 1: Measures of goal characteristics.

usefulness	interestingness	structure	attainability
long-term utility	value congruence, importance	hierarchy-low level	plannability
relative utility	vitality, self-congruence	hierarchy-high level	immediate actionability

Procedure. All participants gave informed consent. The experiment began by asking the participant "What are you up to this week?" and "Which of those goals is most important to you?". The participant then rated the selected goal on the selected GCQ subscales and the self-concordance scale. Next, all participants completed the propensity to plan scale. Participants in the experimental or active control condition then completed the instructions and quiz on our web application and proceeded to our web application. We required them to specify at least five candidate goals with at least two milestones each, along with estimates of the subjective values of achieving each goal and time estimates for the goals and milestones. Participants in those conditions then proceeded to rate the goal our web application recommended them to prioritize first on the GCQ subscales and the self-concordance scale. Participants in the passive control condition were asked to list their goals for next week and to rate the most important one on the GCQ subscales and the self-concordance scale. All participants then rated their likelihood and motivation to pursue the goal and reported how much time they planned to devote to it next week. Participants in the experimental and the active control condition then completed the meCue survey and rated how likely they would recommend our application to a friend. Lastly, all participants were asked to report their age and gender.

Selection of statistical tests. To select appropriate statistical tests, we inspected the histograms of each outcome variable for each group. If an outcome variable was approximately Normally distributed in all conditions, then we selected parametric tests. In the case of the GCQ scores, the distribution was left-skewed. We therefore applied the transformation $f(x) = \sqrt{x_{\max} - x}$, where $x_{\max}$ is the highest possible score. The statistical analyses were performed on the transformed data. The descriptive statistics are reported for the original data. For all other variables whose distribution substantially deviated from a Normal distribution, we selected appropriate nonparamtric tests, as documented below.

3.2 Results

As detailed below, we found that the intuitively compelling approach of suggesting the goal that would allow the user to earn the largest number of points per hour was harmful rather than helpful.

3.2.1 How good were the AI-generated goal suggestions?

User satisfaction with the goal suggestions. On average, the participants in the experimental condition rated their satisfaction with the goal suggestions of our digital assistant as 0.61 on a scale from −3 to 3 (median: 0.88, SD: 1.18) which is significantly greater than zero ($t(82) = 4.73$, $p < .001$, $d = 0.52$). However, users were no more satisfied with the optimal goal suggestions than with the randomly generated goal suggestions (M=0.57, SD=1.24, F(1,162)=0.31, p=.58, d=0.03).

Characteristics of the suggested goals. Contrary to Hypotheses 1, using our digital goal-setting assistant led to goals that participants rated as less useful than goals produced by unassisted goal-setting ($F(1, 323) = 21.22$, $p < .0001$, $\eta^2 = 0.067$) and randomly selected goals ($F(1, 323) = 9.62$, $p = .0021$, $\eta^2 = 0.027$). Contrary to Hypotheses 2, using our digital goal-setting assistant led to goals that people rate as less interesting than unassisted goal-setting ($F(1, 647) = 24.17, p < .0001$, $\eta^2 = 0.030$) or selecting goals at random ($F(1, 651) = 15.25, p = .0001$, $\eta^2 = 0.019$). Moreover, this effect was weaker for people who engage in less short-term planning ($F(1, 651) = 6.31, p = .0123$, $\eta^2 = 0.008$). Contrary to Hypotheses 3, the goals suggested by our digital assistant are less well integrated into the person's goal hierarchy than goals that they set for themselves ($F(1, 323) = 7.13, p = .0080$, $\eta^2 = 0.020$) and randomly selected goals ($F(325, 1) = 8.7, p = .0034$, $\eta^2 = 0.024$). Moreover, we found that the perceived attainability was higher for the goals recommended by our digital assistant than the goals people set for themselves ($F(1, 323) = 0.98, p = .0078$, $\eta^2 = 0.017$) but did not differ between the goals selected by our algorithm and randomly selected goals ($F(1, 325) = 1.60, p = .21$, $\eta^2 = 0.003$).

Motivation. Participants in the passive control condition felt significantly more motivated to pursue the goal that they had chosen for themselves (median: 8/10) than the participants in the experimental condition felt motivated to pursue the goal that our application suggested to them (median: 6.3/10; $Z = -5.78, p < .001$, $r = -0.45$). This motivation did not differ between the experimental and the active control condition (median: 6/10; $Z = 0.47, p = .63$, $r = 0.04$). In addition, the self-concordance score was significantly higher for the passive control condition (median: 4) than for the experimental condition (median: 2; $Z = -2.55, p = .01, r = -0.20$). Moreover, participants in the experimental condition intended to devote significantly less time to the pursuit of the suggested goal (median: 2h, mean: 3.5h, SD: 3.4h) than participants in the passive control condition intended to devote to their self-chosen goal (median: 8h, mean: 16.2h, SD: 19.9h; $Z = -6.38, p < .0001, z = -.050$).

3.2.2 User Experience ratings of the digital goal-setting assistant.

meCue Questionnaire. Our application's median overall user experience rating was +1.5 on a scale from -5 to +5 (M=0.8, SD=2.6), confirming that the overall reception of our digital goal-setting assistant was significantly positive ($t(82) = 2.59$, $p = .01$, $d = 0.28$, 95% CI: $[0.18, 1.35]$). The average ratings of our application's usefulness (4.7/7) and usability (5.2/7) were also favorable. The means and standard deviations of all meCue dimensions are shown in Table 2. Contrary to H4, whether our application provided AI-generated goal suggestions or randomly selected goal suggestions had no effect on the users' overall satisfaction ($F(1, 162) = 0.63$, $p = .43$, $\eta^2 = 0.002$) or their assessments of the application's usefulness ($F(1, 325) = 1.3, p = .2746, \eta^2 = 0.003$; H4a). Moreover, the experimental group experienced less positive emotion and more negative emotion than the active control group ($F(3, 651) = 3.17, p = .0240, \eta^2 = 0.006$; H4b). Participants who received the purportedly optimal goal suggestions even reported a significantly lower intent to use our application in the future than participants who received randomly selected goal suggestions ($F(1.325) = 5.31, p = .02, \eta^2 = 0.016$; H4c). Participants with a higher propensity for long-term planning were more inclined to use our application again in the future ($F(1, 325) = 5.9, p = .0157, \eta^2 = 0.018$).

Feedback and suggestions. On average, participants agreed that they would prefer fewer goals ($t(162) = 7.28, p < .001, d = 0.57$) rather than more goals ($t(162) = -8.52, p < .0001, d = -0.67$). They disagreed with the statement that they would not know how to

Table 2: Mean ± SD of the meCue user experience ratings for our first prototype (SD: Standard deviation)

usefulness	usability	aesthetics	status	commitment	pos. emotion	neg. emotion	loyalty	intent to use
4.3 ± 1.4	5.2 ± 1.3	3.3 ± 1.4	2.9 ± 1.2	2.0 ± 1.1	2.8 ± 1.1	3.2 ± 1.2	2.8 ± 1.2	2.7 ± 1.2

pursue the proposed goals ($t(162) = -9.83, p < .0001, d = -0.77$) and agreed that they would be able to achieve them without a to-do list app ($t(162) = 6.43, p < .0001, d = 0.50$) or with a to-do list app ($t(162) = 7.71, p < .0001, d = 0.60$). Their ratings indicated they did not think a to-do list app could improve their ability to achieve the recommended goals ($t(162) = 0.24, p = .81, d = 0.02$).

4 DISCUSSION

Our findings demonstrated that AI algorithms can select goals with specific characteristics. However, according to the users, the characteristics that our algorithm selected for were not the ones we had intended. Thus, either our algorithm's recommendations were counterproductive, or people's intuitive judgments are biased against goals that are objectively good, or both. Although our participants may have been partial to goals they chose for themselves, this cannot explain why they favored the randomly selected goals over those selected by our algorithm. Either way, our findings suggest that aligning the objective function of the goal suggestion algorithm with what is best for the user is a challenging problem. This echoes current concerns about the challenges of aligning the goals of AI with human values [4, 13, 20].

Assuming that our participants' ratings of their goals are trustworthy, our findings highlight two conceptual problems we did not anticipate. First, although we intended to direct people toward important far-sighted goals, the goals with the highest number of points/hour were more actionable but less farsighted, less useful, and less interesting than randomly selected goals and the user's previous goals. This suggests that – in practice – the points/hour metric was usually dominated by the user's estimate of how little time it will take to achieve the goal rather than the goal's value. That is, our digital assistant failed to achieve the objective of recommending the most important goals. Instead, it recommended small goals that could be completed quickly. The root of this problem might be that users underestimated how much more valuable their most important goals are than their smallest goals. Our application might have exacerbated people's reduced sensitivity for differences between large values [8] by asking them to estimate all goals' values on a compressed scale from 1% to 100%. Moreover, users likely underestimated the time required to achieve some seemingly smaller goals [12]. Future work should investigate these issues and devise ways to help users make more accurate estimates, correct systematic errors, or change the criterion for selecting goal recommendations. For instance, the user's time estimates could be improved by providing data on how long it took others to achieve similar goals in the past [12]. More generally, designing user interfaces that elicit accurate information about people's goals could be crucial for successful AI alignment. Second, consistent with self-determination theory [6], users were more motivated to pursue goals they chose for themselves than those suggested. This makes presenting goal suggestions in such a way that they maximally improve goal-setting without undermining people's autonomy a critical design challenge for future work. A simple improvement could be asking the user whether they want to prioritize the top suggestion or select one of the other suggestions instead.

One limitation of our study was that we did not ask participants to explain why they thought the suggested goal was (un)helpful. Moreover, future work should complement our subjective measures of goal quality with objective measures, such as the effect of the goal suggestions on subsequent productivity. Also, while the present study took a top-down approach to designing the goal-setting assistant, future work should be grounded in the user's needs, goals, and projects. Relatedly, while we evaluated the goal-setting assistant with a convenience sample, it might be beneficial to identify a specific target group and design the goal-setting assistant around their needs. People's attitudes towards and current use of goal-setting apps might be relevant. In addition, we identified the propensity for long-term planning as a promising predictor of people's intention to use our application. One exciting target group could be technology-affine university students looking for life/career goals that will maximize their positive social impact in the long run. Interviews could inform the design of a suitable app by providing insights into the problem of choosing altruistic long-term goals, the kinds of goals students consider, and how they think about time and value in that context.

ACKNOWLEDGMENTS

This research was supported by grant number 1757269 from the National Science Foundation.

REFERENCES

[1] Yossi Aviv and Amit Pazgal. 2005. A partially observed Markov decision process for dynamic pricing. *Management science* 51, 9 (2005), 1400–1416. https://doi.org/10.1287/mnsc.1050.0393

[2] Laura Aymerich-Franch and Iliana Ferrer. 2022. Investigating the use of speech-based conversational agents for life coaching. *International Journal of Human-Computer Studies* 159 (2022), 102745. https://doi.org/10.1016/j.ijhcs.2021.102745

[3] Shalabh Bhatnagar, Emmanuel Fernández-Gaucherand, Michael C Fu, Ying He, and Steven I Marcus. 1999. A Markov decision process model for capacity expansion and allocation. In *Proceedings of the 38th IEEE Conference on Decision and Control (Cat. No. 99CH36304)*, Vol. 2. IEEE, IEEE, Phoenix, AZ, USA, 1380–1385. https://doi.org/10.1109/CDC.1999.830146

[4] Brian Christian. 2020. *The alignment problem: Machine learning and human values.* WW Norton & Company, New York, USA.

[5] Saksham Consul, Jugoslav Stojcheski, Valkyrie Felso, and Falk Lieder. 2021. *Optimal To-Do List Gamification for Long Term Planning.* Technical Report. arXiv. 30 pages.

[6] Edward L Deci and Richard M Ryan. 2012. Self-determination theory. In *Handbook of theories of social psychology*, P. A. M. Van Lange, A. W. Kruglanski, and E. T. Higgins (Eds.). Sage Publications Ltd, London, UK, 416–436. https://doi.org/10.4135/9781446249215.n21

[7] Jiahui Du, Weijiao Huang, and Khe Foon Hew. 2021. Supporting students goal setting process using chatbot: implementation in a fully online course. In *2021 IEEE International Conference on Engineering, Technology & Education (TALE)*. IEEE, IEEE, Wuhan, China, 35–41. https://doi.org/10.1109/TALE52509.2021.9678564

[8] David Fetherstonhaugh, Paul Slovic, Stephen Johnson, and James Friedrich. 1997. Insensitivity to the value of human life: A study of psychophysical numbing. *Journal of Risk and uncertainty* 14, 3 (1997), 283–300. https://doi.org/10.1023/A:1007744326393

[9] Adam Maria Gadomski, Sandro Bologna, Giovanni Di Costanzo, Anna Perini, and Marco Schaerf. 2001. Towards intelligent decision support systems for emergency managers: the IDA approach. *International Journal of Risk Assessment and Management* 2, 3-4 (2001), 224–242. https://doi.org/10.1504/IJRAM.2001.001507

[10] Hernan Gonzalez Cruz, Mike Prentice, and Falk Lieder. 2021. 'What Do You Want in Life and How Can You Get There?' An Evaluation of a Hierarchical Goal-Setting Chatbot. Abstract of presentation at the 13th SSM Virtual Congress. In *13th Annual meeting of the Society for the Science of Motivation*. Society for the Science of Motivation, Virtual Congress, 3.

[11] Gabriela Yukari Iwama, Felix Weber, Mike Prentice, and Falk Lieder. 2021. *Development and Validation of a Goal Characteristics Questionnaire.* Technical Report. OSF Preprints. 61 pages.

[12] D. Kahneman and A. Tversky. 1979. Intuitive prediction: biases and corrective procedures. *TIMS Studies in Management Science* 12 (1979), 313–327. Issue 12.

[13] Sebastian Krakowski. 2021. Artificial Intelligence and Business Ethics: Goal Setting and Value Alignment as Management Concerns. In *Academy of Management Proceedings*, Vol. 2021. Academy of Management Briarcliff Manor, NY 10510, Academy of Management, New York, USA, 14636. https://doi.org/10.5465/AMBPP.2021.14636abstract

[14] Falk Lieder, Owen X Chen, Paul M Krueger, and Thomas L Griffiths. 2019. Cognitive prostheses for goal achievement. *Nature human behaviour* 3, 10 (2019), 1096–1106. https://doi.org/10.1038/s41562-019-0672-9

[15] Falk Lieder and Tom Griffiths. 2016. Helping people make better decisions using optimal gamification.. In *CogSci 2016*. Cognitive Science Society, Philadelphia, Pennsylvania, 2075–2080.

[16] Brian R Little, Katariina Salmela-Aro, and Susan D Phillips. 2017. *Personal project pursuit: Goals, action, and human flourishing.* Psychology Press, Washington DC, USA.

[17] John G Lynch Jr, Richard G Netemeyer, Stephen A Spiller, and Alessandra Zammit. 2010. A generalizable scale of propensity to plan: The long and the short of planning for time and for money. *Journal of consumer research* 37, 1 (2010), 108–128. https://doi.org/10.1086/649907

[18] Michael Minge, Manfred Thüring, Ingmar Wagner, and Carina V Kuhr. 2017. The meCUE questionnaire: a modular tool for measuring user experience. In *Advances in Ergonomics Modeling, Usability & Special Populations*, M. Soares, C. Falcao, and T. Z. Ahram (Eds.). Springer, Florida, USA, 115–128. https://doi.org/10.1007/978-3-319-41685-4_11

[19] Luiz Guilherme Nadal Nunes, Solon Venancio de Carvalho, and Rita de Cássia Meneses Rodrigues. 2009. Markov decision process applied to the control of hospital elective admissions. *Artificial intelligence in medicine* 47, 2 (2009), 159–171. https://doi.org/10.1016/j.artmed.2009.07.003

[20] Stuart Russell. 2019. *Human compatible: Artificial intelligence and the problem of control*. Penguin, New York, USA.

[21] Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, et al. 2020. Mastering atari, go, chess and shogi by planning with a learned model. *Nature* 588, 7839 (2020), 604–609. https://doi.org/10.1038/s41586-020-03051-4

[22] Kennon M Sheldon, Andrew J Elliot, Richard M Ryan, Valery Chirkov, Youngmee Kim, Cindy Wu, Meliksah Demir, and Zhigang Sun. 2004. Self-concordance and subjective well-being in four cultures. *Journal of cross-cultural psychology* 35, 2 (2004), 209–223. https://doi.org/10.1177/0022022103262245

[23] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. 2017. Mastering the game of go without human knowledge. *nature* 550, 7676 (2017), 354–359. https://doi.org/10.1038/nature24270

[24] Haili Song, C-C Liu, Jacques Lawarrée, and Robert W Dahlgren. 2000. Optimal electricity supply bidding by Markov decision process. *IEEE transactions on power systems* 15, 2 (2000), 618–624. https://doi.org/10.1109/59.867150

[25] Jugoslav Stojcheski, Valkyrie Felso, and Falk Lieder. 2020. *Optimal to-do list gamification.* Technical Report. arXiv. 29 pages.