Tools for (Changing) Thought: Scaffolding Cognitive Skills Development within Mental Health Contexts

Petr Slovak
petr.slovak@kcl.ac.uk
King's College London
London, UK

Overview

This paper builds on two prior award-winning CHI-submissions—one from last year [4], and one to be presented at CHI'25 (see here)—which together exemplify a conceptual and design approach to supporting cognitive skills development with generative AI systems. I would be excited to discuss these ideas in more depth at the workshop, and in particular explore the extent to which these could be expanded beyond mental health interventions – especially, e.g., around work-based interaction support and cognitive skills development.

In what follows, I will **first briefly outline the upcoming CHI'25 submission**, where we used an innovative human-AI collaboration workflow to enable people to articulate and share stories of their experience regardless of their writing ability. The gen-AI system was designed to reduce the cognitive load of narrative creation while retaining the participants' own words in the resulting stories. Since the CHI paper submission, we've been already able to start deploying the tool in range of different contexts, spanning both simple data collection as well as intervention development. As an example, this included a KCL / Cambridge / Stanford collaboration to collect a dataset of 950 young people's stories of their most challenging experiences on social media and the types of support they would have preferred to receive – a type of data that is, bizarrely, so far missing from the online social media literature, certainly at such a scale. The use of the tool also elicited surprisingly high engagement from the teens: e.g., over 75% of youth reporting the tool as 'very' or 'extremely' helpful for articulating their experience.

Second, I will explain **how the micro-narratives emerged from a broader design framework outlined in our CHI'24 paper**, co-authored with Sean Munson. Our framework aims to bridge the tension between designing for *psychological efficacy* (relying on prior work) and *design innovation* (aiming to disrupt and change status quo). In particular, we show how theories of change taken from psychological literatures can be thought-off as *(cognitive) trajectories of experience*, and thus translated into traditional design briefs in ways to match well with traditional HCI design methods. We further argue how such approach can help rethink how mental health interventions are designed and developed, supporting technology-enabled innovation within mental health. As illustrated by the micro-narratives work, this approach seems to be particularly well suited for developing gen-AI enabled 'tools

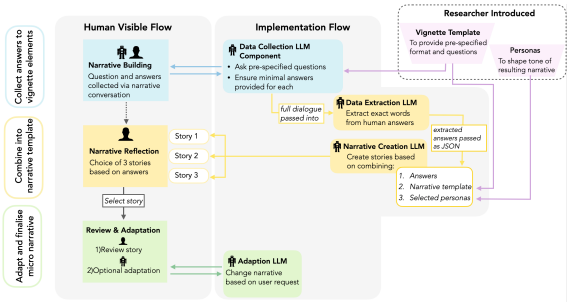

Figure 1: Overview of the three-stage human AI workflow

for thought': it provides one possible approach of incorporating existing psychological theories into design (as temporal cognitive processes described by theories-of-change) while also enabling substantive innovation in how such theories of change can be designed for and delivered.

Micro-narratives: A Scalable Method for Eliciting Stories of People's Lived Experience

Motivation and design goals. This work was motivated by the challenges of collecting rich, qualitative narratives of participants experiences at scale (e.g., >150 participants) without undue participant burden and excessive costs of conducting the research: (i) interviews or detailed diary studies provide the necessary depth of understanding of participants' experience, but are resource intensive and pose a substantial burden on the populations and the research team; (ii) approaches such as cultural probes are often bespoke and are similarly complicated to scale and resource (cf., [2]); and (iii) while questionnaires (including EMAs) can be deployed at scale, they often struggle to capture the depth of users' emotional experiences (cf., [6]).

Our design aimed to explore the opportunities for digitally mediated support that would, at the same time, provide: (1) enough *open-endedness*, to capture participants' lived experience, in their words; (2) enough *consistency* in the set of core aspects covered within the stories, so that specific research questions can be addressed; while (3) *reducing the burden* for participants, so they remain willing to create and share their narratives, and ideally find value for themselves in doing so.

Designing for cognitive trajectories. Our design process directly drew on the design framework [5] outlined in the next section – with the focus on clearly articulating the users' *cognitive trajectory* that the design is aiming to support. In this case, we considered

Tools for Thought worskhop, Yokohama, Japan

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

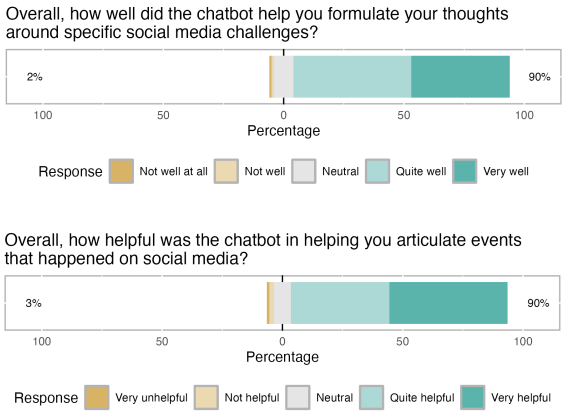

Figure 2: Acceptability feedback from initial pilot

how the mental process of accomplishing this task might look for the participants if they were to be asked to articulate similar template-based narratives without any AI-support; and, most importantly, *which of the steps in such a flow might be most difficult and/or burdensome.*

The resulting system is based on a combination of *deterministic LLM-chaining techniques* to provide a *cognitive scaffolding* for the participants' narrative creation process. In other words, we have brought in a particular theory of narrative creation, and used this psychological theory to decompose the task into steps that:

- are easy for individuals to accomplish (which humans do),
- are difficult for humans but can be automated (which AI does, based on directives from researchers),
- while giving individuals agency to build and revise the resulting narratives to make them their own.

Figure 1 provides an overview of the agent-like decomposition of gen-AI calls, which was directly mapped onto the cognitive flow of the participants' tasks that we were aiming to support. From the psychological perspective, it is based on a re-thinking of how 'vignettes' are used in psychology to date – from acting as a storytelling device from researchers-to-participants to a format that enables storytelling from participants-to-researchers. More specifically, the systems tested the assumption that articulation of a story might consist of collecting a set of carefully selected 'fragments' of the situation (which are individually easy to answer), but which can be then combined into a coherent narrative (based on an underlying 'vignette' template) and serve as a helpful starting point for further adaptation if needed.

Findings and ongoing impact. The CHI paper reports on three studies within the context of a proof-of-concept system aimed at collecting stories of youth's difficult experiences on social media: an initial pilot (N=100), a 2-week asynchronous remote community co-design (N=30), and an experimental study (N=254 youth) that compared the prototype with an analogously worded open-ended survey question (as the closest comparator) with a study design that allowed us to compare both between- and within-subject effects.

Both **qualitative and quantitative results** were highly positive: when comparing to the traditional open-text survey questions, youth were, 4.5x more likely to report that micro-narrative was easier to capture their experience, 2.8x likely that it was more appropriate for (other) youth, 6x more likely that is it more helpful in making sense of the experience, and 4.8x more likely to better support them to think about how to address their social media challenge. Moreover, we also **mapped out a range of potential use-cases for micro-narrative tools**, based on informal discussions with 4 C-level staff at major national and international non-profits, as well as 14 established researchers (median citations 13.1k) across a range of disciplines, including clinical psychology, behavioural health, communication studies, implementation science, and HCI. This resulted in the dataset of 950 youths' stories mentioned in the introduction as well as ~10 ongoing active deployments of the tool since September'24.

Implications for workshop discussions. I see the computational and psychological infrastructure behind micro-narratives as one possible example of a 'blueprint' for developing and deploying 'tools-for-thought' – i.e., an approach whereby gen-AI components enable participants to accomplish a task by support specific aspects of the underlying cognitive flow, while deeply grounded in underlying psychological theories. Moreover, the *design process* leading up to the micro-narratives could be potentially be applicable to a range of other psychological constructs (and underlying cognitive trajectories) – as per the design and innovation framework I turn to in the next section.

Applications of a Modular Framework to Guide Psychosocial Intervention Design

This workshop is interested in exploring how gen-AI systems can act as 'functional extensions of human cognition', with focus on both how the current human-AI interactions affect human cognition, and how we could design to 'augment and protect' it. In both cases, the research questions are grounded around the impact of a (socio-)technical system on users' thought patterns or skills—potentially leading to lasting change, whether this is positive (e.g., increase in critical thinking or knowledge) or negative (e.g., over-reliance on AI leading to cognitive skills atrophy).

In the rest of this paper, my main argument is that such research questions—even if framed in the context of work or education—might benefit from the methodological and design approaches originally developed within the context of mental health interventions. I would suggest that such overlaps will be particularly strong in contexts such as education support (e.g., coaching / tutoring) and meta-cognition (e.g., supporting users in developing deeper reflection, sense-making, or critical thinking skills) – where the core design target is a shift in a cognitive skill as part of the users' interaction with the system.

Framework overview: In what follows, I briefly outline a design framework, which was originally designed to help HCI researchers and psychologists working in the contexts of mental health interventions to gain *common language and a conceptual model* in ways that include the 'psychological-active' elements into the design process, and offer an interface to communicate our findings to our colleagues in the mental health space (and vice-versa).

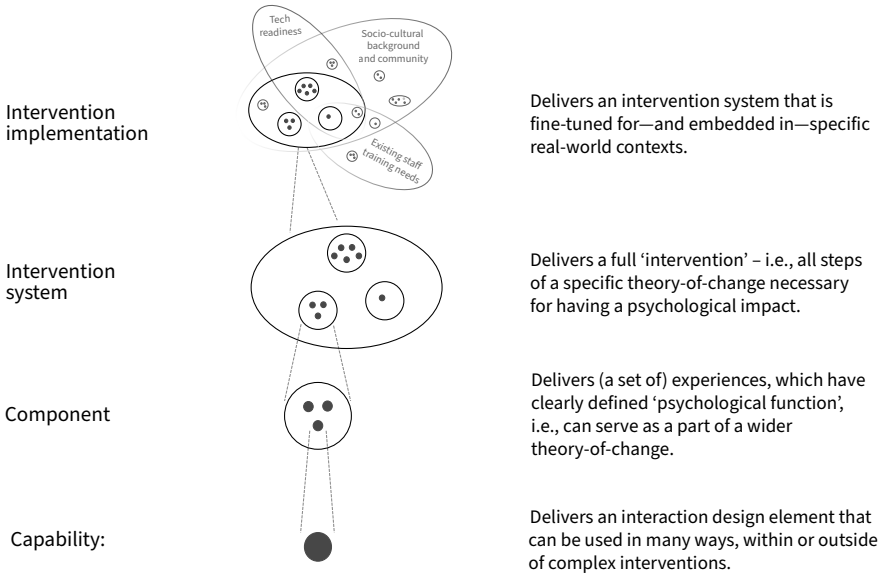

Figure 3: Our framework argues that psycho-social interventions can be seen prescribing 'sets of experiences' that are expected to lead to psychological effects for the participants ...and thus can be translated into design briefs. This figure provides an overview of the four key types of design briefs—at the capability, component, intervention system, and intervention implementation level—that correspond to different functions within an intervention.

My suggestion is that the framework could be helpful in the context of this workshop in two ways: First, by *highlighting the conceptual and analytical approaches that can be used to analyse the effects of novel systems ('interventions') on human thought process* – in particular, the focus on 'theories of change' as the methodological tool to understand the psychological mechanims that lead to any impacts on cognition. I note that while psychological interventions usually start a with given theory-of-change and develop a digital system (i.e., the intervention) to deliver such content, it is also possible to use the theory-of-change analysis to understand the reasons behind impacts of existing systems (e.g., analysing why a gen-AI system negatively impact users' decision making skills). Second, by *providing a design framework that can combine psychological knowledge* (e.g., how to support 'good' metacognition) *with the design process* (i.e., how these psychological mechanisms could be actually delivered); and to do so in ways that empower the interdisciplinary collaboration necessary.

While we cannot fully outline the argument here, the high-level overview is below – I encourage readers to read the full paper if of interest.

Theories-of-change as (cognitive) experience trajectories. First, we propose that a useful model for HCI designers is to see the theories-of-change underpinning mental health interventions as prescribing a **particular set of experience trajectories** (cf., [1]) that *are expected to lead to the psychological effects for the participant(s).* In particular, we show how this framing can support designers and researchers in fluently translating between interventions' **theories of change** (i.e., descriptions of the psycho-active elements and how these bring about the change in mental health) and **design briefs** (as an established way of specifying interaction design aims in HCI). In the context of gen-AI impacts, it is also possible to see this process reversed – by observing impacts of an existing system, we can start inferring the possible experience trajectories (and underlying theories of change) that lead to these outcomes.

In both ways, such re-framing of *intervention design* into *experience design* enables us to bridge the tension between psychological efficacy and design impact – both now are positioned as complementary requirements / boundaries on the experience to be supported. In our experience, such framing is then understandable for psychologists and HCI researchers alike.

Modularity of theories of change. Second, we argue that approaching theories of change as modular design briefs *gives technologists and designers flexibility to imagine new technical and interface capabilities* that can implement parts of the functionality prescribed by theory of change; as well as propose new theory of change opportunities based on emerging technologies. To illustrate this reasoning, we proposed a vocabulary of four conceptual types of design briefs—capabilities, components, intervention systems, and intervention implementations—corresponding to different functional 'levels' within a theory of change where design contributions can happen, and the associated evaluation methodologies required – see Figure 3 and the original paper for details.

In the context of gen-AI systems in work or educations contexts, this conceptual structure can still help unpack the individual components / design decisions that are likely to drive positive (or negative) impacts on human cognition, as well as support design to amplify (or counterbalance) such effects. For example, similar approach has been key for the design of the micro-narrative tool described in the

previous section. In particular, it has helped us decompose the cognitive task into the individual components (that together combine into a full narrative flow), as well as isolate the design iteration for each step of the process. It is also guiding the further extension of the system, e.g., as we are expanding the micro-narrative flow to act as a meta-cognitive intervention (such as through empowering reflection, decentering, and sense-making processes) rather than 'just' a data collection tool.

Implications for workshop discussions. The original goal of the framework was to articulate the different ways in which HCI can do influential innovation work and evaluate its success on foundational levels (such as new capabilities or components) without the requirements to immediately embed these into higher levels (systems or implementations). This can help sidestep the methodological difficulties with evaluation models that are not well aligned with traditional design practice in HCI (such as large scale, multi-site randomised controlled trials) but that are crucial for evaluating the psychological impact of intervention systems / implementations, and the corresponding uptake of designs in mental health community (cf., [3]).

As the core dilemma of understanding the impact of technical innovation in the context of psychological theories is likely to exist in any setting where the goal is to shift users' cognitive process in novel ways, my expectation is that many aspects of the framework could be directly extended into the contexts explored within this workshop – and I would be delighted to have a chance to explore such connections within the workshop discussions.

Final thoughts

It was interesting for me to engage with the 'topics of interest' questions posed by the organisers – especially as we are already grappling with many of these directions around the aspects of 'human thinking / learning' within mental health space. A unifying characteristic of the mental health interventions design is that the ultimate aim to impact participants *skills and thought processes*: using the system should alter something about how they perceive and engage with their reality. In other words, it is rarely the 'task'—or its output—that is the actual goal, but rather the impact that going through such a task has on the participants' cognition, ideally in ways that sustain such changes beyond the interaction itself.

Such focus will be naturally shared also within any *learning* system. I am however curious to understand the extent to which such models can be helpful in the domains of *work* and *creativity*, which are the other foci of this workshop, and for which many of the questions around genAI impacts to workflows and human cognition take a very different perspective.

References

[1] Steve Benford, Gabriella Giannachi, Boriana Koleva, and Tom Rodden. 2009. From interaction to trajectories: designing coherent journeys through user experiences. In *CHI'09*. ACM, 709–718. http://portal.acm.org/citation.cfm?id=1518701.1518812

[2] Seray B Ibrahim, Alissa N. Antle, Julie A. Kientz, Graham Pullin, and Petr Slovak. 2024. A Systematic Review of the Probes Method in Research with Children and Families. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference* (Delft, Netherlands) *(IDC '24)*. Association for Computing Machinery, New York, NY, USA, 157–172. doi:10.1145/3628516.3655814

[3] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. 2011. How to evaluate technologies for health behavior change in HCI research. In *CHI '11*. ACM Press, 3063–3072. doi:10.1145/1978942.1979396

[4] Petr Slovak and Sean A. Munson. 2024. HCI Contributions in Mental Health: A Modular Framework to Guide Psychosocial Intervention Design. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 692, 21 pages. doi:10.1145/3613904.3642624

[5] Petr Slovak and Sean A. Munson. 2024. HCI Contributions in Mental Health: A Modular Framework to Guide Psychosocial Intervention Design. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 692, 21 pages. doi:10.1145/3613904.3642624

[6] Jing Wei, Sungdong Kim, Hyunhoon Jung, and Young-Ho Kim. 2024. Leveraging Large Language Models to Power Chatbots for Collecting User Self-Reported Data. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–35. doi:10.1145/3637364 arXiv:2301.05843