



# Harnessing AI to Optimize Thought Records and Facilitate Cognitive Restructuring in Smartphone CBT: An Exploratory Study

Toshi A. Furukawa<sup>1</sup> · Susumu Iwata<sup>2</sup> · Masaru Horikoshi<sup>1,3</sup> · Masatsugu Sakata<sup>1</sup> · Rie Toyomoto<sup>1</sup> · Yan Luo<sup>1,4</sup> · Aran Tajika<sup>1</sup> · Noriko Kudo<sup>5</sup> · Eiji Aramaki<sup>5</sup>

Accepted: 30 June 2023 / Published online: 7 July 2023  
© The Author(s) 2023

## Abstract

**Background** Effective cognitive restructuring (CR) requires identification of automatic thoughts that underlie experienced emotions. However, accurate recording of thoughts and emotions is challenging when CR is provided in internet cognitive-behavior therapy (iCBT). This study investigated the potential use of the artificial intelligence (AI) including the natural language processing (NLP) to facilitate CR offered in iCBT.

**Methods** We applied the Japanese Text-to-Text Transfer Transformer (T5), one of the most advanced Large Language Models for the NLP, to records of thought-feeling pairs provided by participants in two randomized controlled trials of iCBT. We conducted threefold cross-validated prediction of self-reported feelings based on recorded thoughts. We examined the validity of the predictions by checking them against the human expert judgments and by the efficacy when the thought records were subjected to CR.

**Results** 1626 participants provided 4369 thought-feeling records. The overall prediction accuracy was 73.5%. The self-reported feelings matched the human expert judgments more frequently when they were correctly predicted by the T5 than not (90% vs 37.5%, 95%CI of difference: 34.8 to 70.2%). When subjected to CR, the correctly predicted thought-feeling pairs led to greater reductions in negative feelings than the incorrectly predicted pairs (− 1.54 vs − 1.43 on a scale of 0 to 5, 95%CI of difference: 0.03 to 0.19).

**Conclusions** A new CR module of an iCBT application can incorporate this model and advise the users to revisit and revise their automatic thoughts to reflect their feelings more accurately. Whether such an iCBT application can ultimately lead to greater reductions in depression is to be examined in a future randomized trial.

**Keywords** Cognitive restructuring · Artificial intelligence · Machine learning · Natural language processing · Large Language Model · Smartphone · Internet CBT

## Introduction

Cognitive-behavior therapy (CBT) is the most researched and established psychotherapy for depression and other emotional disorders (Cuijpers et al., 2021; Hofmann et al., 2012). The growing awareness of the impact of common mental disorders across the world is now calling for psychotherapies that can be delivered at scale (Herrman et al., 2022). Various delivery methods, believed to be more efficient than the traditional face-to-face individual format, have been tried for CBT. Of these, the self-help CBT appears most promising (Cuijpers et al., 2019), and with the rapid popularization of the internet we are witnessing a growing number of internet CBT (iCBT) applications (Karyotaki et al., 2021).

✉ Toshi A. Furukawa  
furukawa@kuhp.kyoto-u.ac.jp

<sup>1</sup> Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

<sup>2</sup> Life2Bits, Inc., Tokyo, Japan

<sup>3</sup> National Center of Neurology and Psychiatry, Kodaira, Japan

<sup>4</sup> Department of Medical Education, Kyoto University Graduate School of Medicine, Kyoto, Japan

<sup>5</sup> Nara Institute of Science and Technology, Ikoma, Japan

While CBT now encompasses a range of cognitive and behavioral intervention skills (Furukawa et al., 2021), cognitive restructuring (CR) remains the central technique deeply rooted in the cognitive model of emotional disorders (Beck et al., 1979). The theory posits that it is not the stressful situations themselves but rather their appraisal which leads to negative emotions. In CR, the patients are encouraged to monitor their experienced emotions and their thoughts amid the stressful situations, and then prompted to challenge their initial, automatic and often dysfunctional thoughts. The prerequisite of CR then is the accuracy of the so-called thought record, in which patients record their emotions and thoughts: Unless the patients can accurately identify the thoughts that underlie their emotions, i.e. that directly lead to their emotions, challenging the thoughts would not easily lead to a reduction in their negative emotions. In the traditional face-to-face CBT sessions, psychotherapists achieve such refinement in identified automatic thoughts through Socratic questionings such as the downward-arrow technique.

In the self-help applications, however, the accuracy of the thought record is largely left to the acumen of the users and no work-up of automatic thoughts has been possible. This difficulty may partly explain the recent findings from the component network meta-analysis of iCBT in which CR was not found to be particularly helpful, in contrast with more directive approaches such as behavioral activation or problem solving (Furukawa et al., 2021). In order to augment the efficacy of CR in iCBT, we need to find ways to help users of iCBT in identifying the underlying automatic thoughts that correspond with their experienced emotions.

Recent advances in the natural language processing (NLP), mainly spurred by the emergence of Large Language Models (LLMs), may bring about a sea-level change in this regard. LLMs learn universal language representations from large volumes of text data using self-supervised learning and transfer this knowledge to more specific applications through fine-tuning (Subramanyam Kalyan et al., 2021; Wang et al., 2022). In this study we used thought records from two previous randomized controlled trials of iCBT and examined if the Text-to-Text Transfer Transformer (T5), one of the most advanced LLMs, can help identify thought records in which users may be having difficulties in accurately identifying the automatic thoughts that underlie their emotions. First we trained the T5 to predict the emotions associated with each automatic thought. Next we tested the validity of these NLP-based predictions by comparing them with the human experts' judgements. We finally tested the validity of our assumption that better matched thought-emotion records enable more effective CR by comparing the results of CR of matched vs non-matched thought records.

If the T5 can identify thoughts that may not correspond with the experienced emotions, the smartphone CBT app can issue prompts to reconsider or refine the thoughts recorded,

which would then enable more effective cognitive restructuring and ultimately lead to a greater reduction in depression through the therapy.

## Methods

### Dataset

We used automatic thought records submitted by participants who had taken part in the two randomized controlled trials of smartphone CBT.

One is the FLATT (Fun to Learn to Act and Think through Technology) trial, which examined the add-on effects of the smartphone CBT app called “Kokoro App” (“Kokoro” means the mind in Japanese) to the pharmacotherapy over the pharmacotherapy alone among 164 patients with treatment-resistant depression (Mantani et al., 2017). “Kokoro App” contains five active components for CBT, namely psychoeducation, self-monitoring, behavioral activation, cognitive restructuring and relapse prevention.

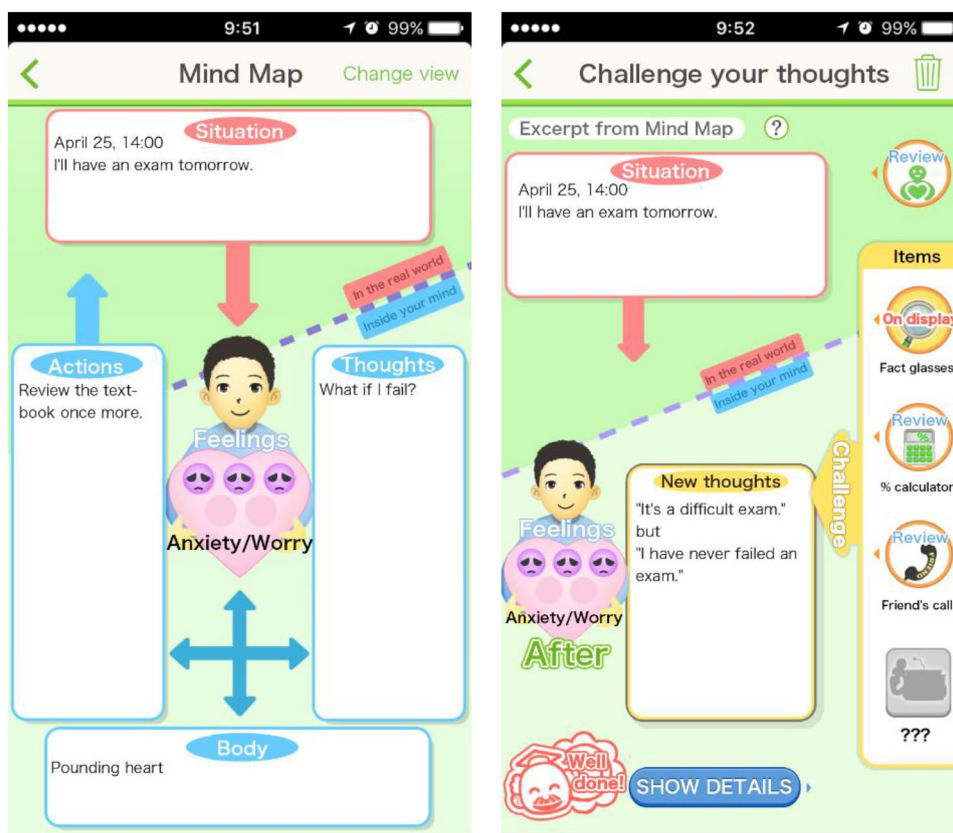
The other is the HCT (Healthy Campus Trial), which used the smartphone CBT app called “Resilience Training App” among 1626 university students, of whom 1093 scored five or higher on the Patient Health Questionnaire-9 (PHQ-9) at baseline (Sakata et al., 2022). “Resilience Training App” is an expanded version of “Kokoro App”, including components for structured problem solving and assertion training, in addition to the original five. The HCT was a fully factorial trial to assess the specific efficacy of the five CBT components of self-monitoring, cognitive restructuring, behavioral activation, problem solving and assertion training. All the participants undertook psychoeducation and relapse prevention but were randomly assigned to presence or absence of the five remaining components, therefore to  $2^5 = 32$  different combinations.

### Variables

In the self-monitoring and cognitive restructuring components, the participants learnt the cognitive model of human responses to stressful situations and filled in “mind maps,” a graphical version of automatic thought records that recorded situations, feelings, thoughts, body reactions and actions. The participants entered free texts into situations, thoughts, body reactions and actions, and chose one of the four basic feelings (sad, anxious, angry or happy) and rated its intensity in six levels of 0–6 (Fig. 1a).

The participants could then use one or more of their mind maps to practice cognitive restructuring. The app provides four items to help the users challenge their automatic thoughts: “Fact glasses,” “% calculator,” “Friend’s call,” and “What now microphone.” (Fig. 1b) “Fact glasses” ask the

**Fig. 1** Screenshot of a mind map and a thought challenge



patients to come up with alternative thoughts based on the reality that do not match the automatic thoughts. “% calculator” asks the users how much they believe in the automatic thought (x %) and then asks what possibilities there are in the remaining (100 – x) %. “Friend’s call” asks a hypothetical question “What advice would you give to your best friend if they told you the very same thought?” And “What now microphone” asks the users “What would be the next best thing you can do if you assumed that the automatic thoughts were true?” The participants can use one or more of these items to develop alternative thoughts and re-evaluate their feelings if they had thought otherwise.

All the participants’ entries into mind maps and thought challenges were automatically uploaded to the remote server and are used in the current analyses.

**Analyses—Prediction**

We used the Japanese Text-to-Text Transfer Transformer (T5) model (<https://huggingface.co/sonoisa/t5-base-japanese>), pretrained on three Japanese corpuses (altogether ca 100 GB) of Wikipedia (<https://ja.wikipedia.org/wiki/%E3%83%A1%E3%82%A4%E3%83%B3%E3%83%9A%E3%83%BC%E3%82%B8>), OSCAR (<https://oscar-corpus.com/>) and

CC-100 (<http://data.statmt.org/cc-100/>), on Python (Version 3.9.0).

We applied the T5 to examine the match between the self-reported feeling and its automatic thought. Assuming that the participants generally correctly identify automatic thoughts, we conducted threefold cross-validation for the model to predict the feelings based on automatic thoughts, using 67% of the sample to train (57%) and fine-tune (10%) the model and the remaining sample to test the model. We then calculated the overall accuracy, precision, recall (sensitivity) and F1-score. Accuracy, precision, recall and F1-score are defined as follows in a 2\*2 confusion matrix.

$$\text{Accuracy} = (TN + TP)/(TN + FP + FN + TP)$$

$$\text{Precision} = TP/(FP + TP)$$

$$\text{Recall} = TP/(FN + TP)$$

$$\text{F1 score} = 2 * \text{Precision} * \text{Recall}/(\text{Precision} + \text{Recall})$$

		Predicted feeling	
		Negative	Positive
Actual feeling	Negative	TN	FP
	Positive	FN	TP

Precision quantifies what proportion of the predicted positive findings are truly positive, recall quantifies what proportion of the truly positive cases can be predicted, and F1 score shows how well the model performs while balancing the two. All these four indexes range between 0 and 1, with the values closer to 1 denoting better performance.

We excluded some of the free-text entries of automatic thoughts that were very brief (five or less Japanese characters in length, which would roughly correspond with two or less words in English). We reasoned that it would be difficult for the T5 to predict the underlying feeling in such cases and that such automatic thoughts were unlikely to be truthful descriptions of the thoughts leading to the feeling. The target feelings were treated as multinomial categories.

## Analyses—Validation

We examined the validity of the model through two tests.

First, we randomly selected 40 thought-feeling pairs each from those in which the T5 predictions were correct and from those in which they were incorrect. We then asked three cognitive behavior therapists (two clinical psychologists (MS and MH) and one psychiatrist (TAF)) to independently guess the feeling behind each of these 80 automatic thoughts. We considered the feelings agreed upon by the independent raters as the gold standard judgements. We hypothesized that the participants' self-reported feelings would be more concordant with the human therapists' judgements as inferred from the automatic thoughts when they were correctly predicted by the T5 than when not.

Second, we hypothesized that the mind maps (thought records), for which the T5 model successfully matched the feeling and thought, would lead to greater reduction in sad,

**Table 1** Baseline characteristics of the FLATT and HCT trials who had contributed one or more mind maps

	FLATT (n = 164)	HCT (n = 1134)
Age, mean (SD)	40.9 (8.9)	21.6 (3.1)
Sex (female), n(%)	87 (53)	645 (57)
PHQ-9 at baseline	13.2 (5.4)	8.2 (3.0)
Total number of mind maps	4369	2813

FLATT Fun to Learn to Act and Think through Technology trial, HCT Healthy Campus Trial, PHQ-9 Patient Health Questionnaire-9

anxious or angry feelings through cognitive restructuring than those for which the T5 failed to match the feeling and the thought. If this hypothesis was confirmed, it would show that the mind maps in which the T5 failed to predict the feelings from the thoughts could benefit from re-assessing the thoughts behind the self-reported feelings.

## Results

### Participants and Data

The 164 participants in the FLATT trial, either in the immediate or in the delayed (waitlist) intervention condition, contributed 4369 mind maps. Of the 1626 participants in the HCT trial, 1134 were assigned to self-monitoring and/or cognitive restructuring components and contributed altogether 2813 mind maps. Table 1 presents the baseline characteristics of these participants. The participants of the FLATT trial were older, more severely depressed at baseline and submitted more mind maps per person than those of the HCT trial. The sex distributions were comparable.

### Prediction

Table 2 presents the confusion matrix between the self-reported feelings and the feelings as predicted by the three-fold application of the T5 model to the feeling-thought pairs. Table 3 shows the precision, recall and F1-score for each feeling. T5 was able to correctly predict the self-reported feeling based on the automatic thought in 5280 (the overall

**Table 2** Confusion matrix between the self-reported vs predicted feelings

		Predicted			
		Sad	Anxious	Angry	Happy
Self-reported	Sad	0.68	0.16	0.08	0.08
	Anxious	0.24	0.67	0.05	0.04
	Angry	0.28	0.14	0.56	0.02
	Happy	0.05	0.03	0.01	0.92

**Table 3** Precision, recall and F1-score

	Precision (or PPV)	Recall (or Sensitivity)	F1-score
Sad	0.65	0.68	0.67
Anxious	0.91	0.90	0.90
Angry	0.69	0.63	0.66
Happy	0.63	0.66	0.64
Average	0.74	0.74	0.74

accuracy of 73.5%) but failed to do so in 1902 (26.5%) cases out of 7182 feeling-thought pairs.

### Validation

We first established the gold standard feeling as estimated from each automatic thought based on the agreed-upon assessment among the three cognitive behavior therapists. The inter-rater reliability of the three independent raters’ assessments was Fleiss’ kappa of 0.74 (95%CI 0.61 to 0.87) for the 40 randomly selected sample of congruent thought-feeling pairs and 0.54 (0.37 to 0.70) for the 40 randomly selected sample from incongruent pairs. We therefore considered the feeling as agreed upon by at least two of the three raters as the gold standard.

The self-reported feelings matched these gold standards in 36 out 40 cases when the T5 algorithm correctly predicted them, but only in 15 out of 40 cases when it could not (90% vs 37.5%, diff = 52.5%, 95%CI of diff = 34.8 to 70.2%,  $p < 0.001$ ).

We next examined the effects of cognitive restructuring for the automatic thoughts when the T5 was able to correctly predict the feeling or when not (Table 4). We limited these analyses to the three negative feelings of sad mood, anxiety or anger. All items tended to produce greater reduction in the negative feelings for the thought records where the T5 successfully matched the self-reported feeling and the thought and taken together the differences were statistically significant (standardized mean difference 0.11, 95%CI 0.03 to 0.19).

### Discussion

We applied the T5, the most advanced NLP using the transformer-based pre-trained language model, to 7182 thought records (thought-feeling pairs) provided by people with major or subthreshold depression. In the threefold internal validation, the T5 was able to correctly predict the self-reported feeling based on the recorded thought in 73.5%

of the thought records. Thought-feeling pairs which the T5 correctly matched showed greater agreement with the gold standard rating as identified by human cognitive-behavior therapists than those pairs which the T5 showed to be discordant. Moreover, when submitted to cognitive restructuring, the former records of thought-feeling pairs led to greater reduction in negative emotions than the latter.

The artificial intelligence (AI) in general and the NLP in particular are finding wider and wider applications across the society, in medicine and in mental health (Wang et al., 2022). A recent review of the use of AI and NPL in mental health has identified five major categories of their usage: to extract clinical symptoms, to classify severity of illnesses, to compare different therapies, to provide psychopathological clues and to challenge the current nosography. For example, an NLP system may aim at detecting suicidal tendencies from the electronic health records, at predicting their severity, at differentiating more vs less adequate deliveries of therapy, or at identifying new psychopathologies based on the use of languages (Le Glaz et al., 2021). Our current study to apply the NLP to rate the quality of the automatic thought records may belong to the third category, namely to distinguish different levels of appropriateness of the therapeutic process.

One of the earlier attempts to apply the NLP to cognitive therapy was a system to classify dysfunctional thoughts into their categories such as “all-or-nothing,” “negative predictions,” “discounting the positive” etc. The researchers collected examples of dysfunctional thoughts from cognitive therapy textbooks and used their classification in the supervised learning. The system seems to have been partially successful but not yet to a clinically applicable degree (Wiemer-Hastings et al., 2004). As the ability of the NLP progresses, more complex systems have been proposed. Burger et al. used the NLP software to classify automatic thoughts provided by crowd-sourced volunteers in terms of their matching schemas as identified by human experts. Among the several models trialed, those based on the recurrent neural networks (RNNs) appeared to perform the best but this study did not provide any clinical applications (Burger et al.,

**Table 4** Reduction in negative emotions after cognitive restructuring

Tools used	Thought records where the T5 failed to match feeling with thought			Thought records where the T5 successfully matched feeling with thought			Standardized mean difference
	Mean	SD	n	Mean	SD	n	
Fact glasses	1.23	1.20	108	1.41	1.17	200	0.15 (− 0.08 to − 0.39)
% calculator	1.38	1.12	270	1.40	1.15	423	0.02 (− 0.13 to 0.17)
Friend’s call	1.50	1.10	321	1.64	1.12	472	0.13 (− 0.01 to 0.27)
What now microphone	1.45	1.10	285	1.63	1.12	443	0.16 (0.01 to 0.31)
Total	1.43	1.12	984	1.54	1.14	1538	0.11 (0.03 to 0.19)

2021). Kawakami et al. developed an NLP-based system to rate the quality of thought records. They trained the RNN to rate the five thought record components (event, thought, mood, behavior, physical symptoms) in accordance with the experts' judgment of their appropriateness. They achieved an accuracy between 0.79 and 0.84 for the five components in the internal cross-validation. The authors are now using this system in their iCBT app to provide feedback to the users when their responses were rated as unlikely to be appropriate by the RNN (Kawakami et al., 2021).

Distinctive from these previous studies, our system was able to demonstrate the discriminatory power of the NLP-based system through external validation. Not only were we able to demonstrate satisfactory discrimination in the internal threefold cross-validation, the mismatching thought-feeling pairs as judged by the NLP system had poorer agreement with the external expert ratings and also led to smaller effectiveness when submitted to cognitive restructuring than those for which the NLP system could correctly predict the thought-feeling matching. These findings suggest that prompting the participants to update/revise the thought that underlie their feelings when the NLP found mismatch between them might be able to lead to more accurate identification of automatic thoughts, to enable more effective cognitive restructuring, and eventually to lead to greater reduction in depression after the treatment.

Some weaknesses of the current study may include the following. First, while showing evidence of internal validity and some external validity, the current study remains essentially exploratory and the ultimate proof of the value of the system is yet to be seen by trialing it in a clinical trial. Second, the NLP is showing very rapid evolution. Although the T5 is currently one of the most advanced NLP systems, newer and better models are being developed. The data to be fed into the NLP are also accumulating as we offer the current CR module to the users. New analyses using a different LLM and an even larger dataset of thought records may improve the performance of the system. However, the model developed in this study already appears promising and ready to be tested out with the real-world users.

We are currently developing a smartphone CR module which incorporates the AI system as demonstrated by the current study. It will provide some feedback to the app users to revisit their automatic thought when there was discordance between the self-reported feeling and the feeling as predicted by the AI system based on the automatic thought. We are currently planning to test out this system in the platform trial of various CBT modules (Furukawa et al., 2023) and to see if such advice can ultimately augment the effect of the iCBT.

**Acknowledgements** The study was supported by a grant-in-aid from the Japan Agency for Medical Research and Development (AMED) (Grant

Number JP22de0107005) to TAF, a JSPS Grant-in-Aid for Scientific Research (Grant Number 21K03049) to MS, a JSPS Grant-in-Aid for Scientific Research (Grant Number 23K02915) to AT, and a JSPS Grant-in-Aid for Scientific Research (Grant Number 23K02935) to RT.

## Declarations

**Competing Interests** TAF reports personal fees from Boehringer-Ingelheim, DT Axis, Kyoto University Original, Shionogi and SONY, and a grant from Shionogi, outside the submitted work; In addition, TAF has patents 2020-548587 and 2022-082495 pending, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe. SI is an employee of Life2Bits, Inc. AT received lecture fees from Sumitomo Dainippon Pharma, Eisai, Janssen Pharmaceutical, Meiji-Seika Pharma, Mitsubishi Tanabe Pharma, Otsuka, and Takeda Pharmaceutical. All the other authors have no conflict of interest to declare.

**Informed Consent** Informed consent was obtained from all individuals who participated in the two trials and provided their thought records.

**Research Involving Animal Rights** No animal studies were carried out for his study.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. Guilford Press.
- Burger, F., Neerincx, M. A., & Brinkman, W. P. (2021). Natural language processing for cognitive therapy: Extracting schemas from thought records. *PLoS ONE*, *16*(10), e0257832. <https://doi.org/10.1371/journal.pone.0257832>
- Cuijpers, P., Noma, H., Karyotaki, E., Cipriani, A., & Furukawa, T. A. (2019). Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: A network meta-analysis. *JAMA Psychiatry*, *76*(7), 700–707. <https://doi.org/10.1001/jamapsychiatry.2019.0268>
- Cuijpers, P., Quero, S., Noma, H., Ciharova, M., Miguel, C., Karyotaki, E., Cipriani, A., Cristea, I. A., & Furukawa, T. A. (2021). Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. *World Psychiatry*, *20*(2), 283–293. <https://doi.org/10.1002/wps.20860>
- Furukawa, T. A., Sukanuma, A., Ostinelli, E. G., Andersson, G., Beevers, C. G., Shumake, J., Berger, T., Boele, F. W., Buntrock, C., Carlbring, P., Choi, I., Christensen, H., Mackinnon, A., Dahne, J., Huibers, M. J. H., Ebert, D. D., Farrer, L., Forand, N. R., Strunk, D. R., ... Cuijpers, P. (2021). Dismantling, optimising, and personalising internet cognitive behavioural therapy for depression: A systematic review and component network meta-analysis

- using individual participant data. *Lancet Psychiatry*, 8(6), 500–511. [https://doi.org/10.1016/S2215-0366\(21\)00077-8](https://doi.org/10.1016/S2215-0366(21)00077-8)
- Furukawa, T. A., Tajika, A., Sakata, M., Luo, Y., Toyomoto, R., Horikoshi, M., Akechi, T., Kawakami, N., Nakayama, T., Kondo, N., Fukuma, S., Noma, H., Christensen, H., Kessler, R. C., Cuijpers, P., & Wason, J. M. S. (2023). Four 2×2 factorial trials of smartphone CBT to reduce subthreshold depression and to prevent new depressive episodes among adults in the community-RESILIENT trial (Resilience Enhancement with Smartphone in LIVING ENvironmentTs): A master protocol. *British Medical Journal Open*, 13(2), e067850. <https://doi.org/10.1136/bmjopen-2022-067850>
- Herrman, H., Patel, V., Kieling, C., Berk, M., Buchweitz, C., Cuijpers, P., Furukawa, T. A., Kessler, R. C., Kohrt, B. A., Maj, M., McGorry, P., Reynolds, C. F., 3rd., Weissman, M. M., Chibanda, D., Dowrick, C., Howard, L. M., Hoven, C. W., Knapp, M., Mayberg, H. S., ... Wolpert, M. (2022). Time for united action on depression: A lancet-world psychiatric association commission. *Lancet*, 399(10328), 957–1022. [https://doi.org/10.1016/S0140-6736\(21\)02141-3](https://doi.org/10.1016/S0140-6736(21)02141-3)
- Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, 36(5), 427–440. <https://doi.org/10.1007/s10608-012-9476-1>
- Karyotaki, E., Efthimiou, O., Miguel, C., BERPohl, F. M. G., Furukawa, T. A., Cuijpers, P., Riper, H., Patel, V., Mira, A., Gemmil, A. W., Yeung, A. S., Lange, A., Williams, A. D., Mackinnon, A., Geraedts, A., van Straten, A., Meyer, B., Björkelund, C., Knaevelsrud, C., ... Forsell, Y. (2021). Internet-based cognitive behavioral therapy for depression: A systematic review and individual patient data network meta-analysis. *JAMA Psychiatry*, 78(4), 361–371. <https://doi.org/10.1001/jamapsychiatry.2020.4364>
- Kawakami, N., Imamura, K., Watanabe, K., Sekiya, Y., Sasaki, N., & Sato, N. (2021). Effectiveness of an internet-based machine-guided stress management program based on cognitive behavioral therapy for improving depression among workers: Protocol for a randomized controlled trial. *JMIR Research Protocols*, 10(9), e30305. <https://doi.org/10.2196/30305>
- Le Glaz, A., Haralambous, Y., Kim-Dufor, D.-H., Lenca, P., Billot, R., Ryan, T. C., Marsh, J., DeVylder, J., Walter, M., Berrouiguet, S., & Lemey, C. (2021). Machine learning and natural language processing in mental health: Systematic review. *Journal of Medical Internet Research*, 23(5), e15708. <https://doi.org/10.2196/15708>
- Mantani, A., Kato, T., Furukawa, T. A., Horikoshi, M., Imai, H., Hiroe, T., Chino, B., Funayama, T., Yonemoto, N., Zhou, Q., & Kawanishi, N. (2017). Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: Randomized controlled trial. *Journal of Medical Internet Research*, 19(11), e373. <https://doi.org/10.2196/jmir.8602>
- Sakata, M., Toyomoto, R., Yoshida, K., Luo, Y., Nakagami, Y., Uwamoto, T., Shimamoto, T., Tajika, A., Suga, H., Ito, H., Sumi, M., Muto, T., Ito, M., Ichikawa, H., Ikegawa, M., Shiraishi, N., Watanabe, T., Sahker, E., Ogawa, Y., ... Furukawa, T. A. (2022). Components of smartphone cognitive-behavioural therapy for subthreshold depression among 1093 university students: A factorial trial. *Evidence-Based Mental Health*, 25(e1), e18–e25. <https://doi.org/10.1136/ebmental-2022-300455>
- Subramanyam Kalyan, K., Rajasekharan, A., & Sangeetha, S. (2021). AMMUS: A survey of transformer-based pretrained models in natural language processing. *arXiv e-Prints*. <https://doi.org/10.48550/arXiv.2108.05542>
- Wang, H., Li, J., Wu, H., Hovy, E., & Sun, Y. (2022). Pre-trained language models and their applications. *Engineering*. <https://doi.org/10.1016/j.eng.2022.04.024>
- Wiemer-Hastings, K., Janit, A. S., Wiemer-Hastings, P. M., Cromer, S., & Kinser, J. (2004). Automatic classification of dysfunctional thoughts: A feasibility test. *Behavior Research Methods Instruments, & Computers*, 36(2), 203–212. <https://doi.org/10.3758/bf03195565>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.