Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

GLoHBCD: A Naturalistic German Dataset for Language of Health Behaviour Change on Online Support Forums

Selina Meyer David Elsweiler

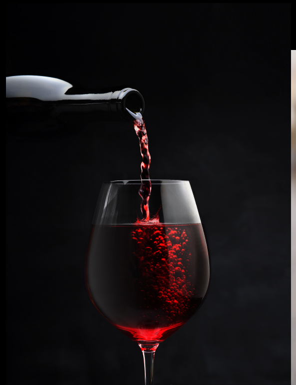
LREC 2022 June 20th-25th, Marseille

Universität Regensburg

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

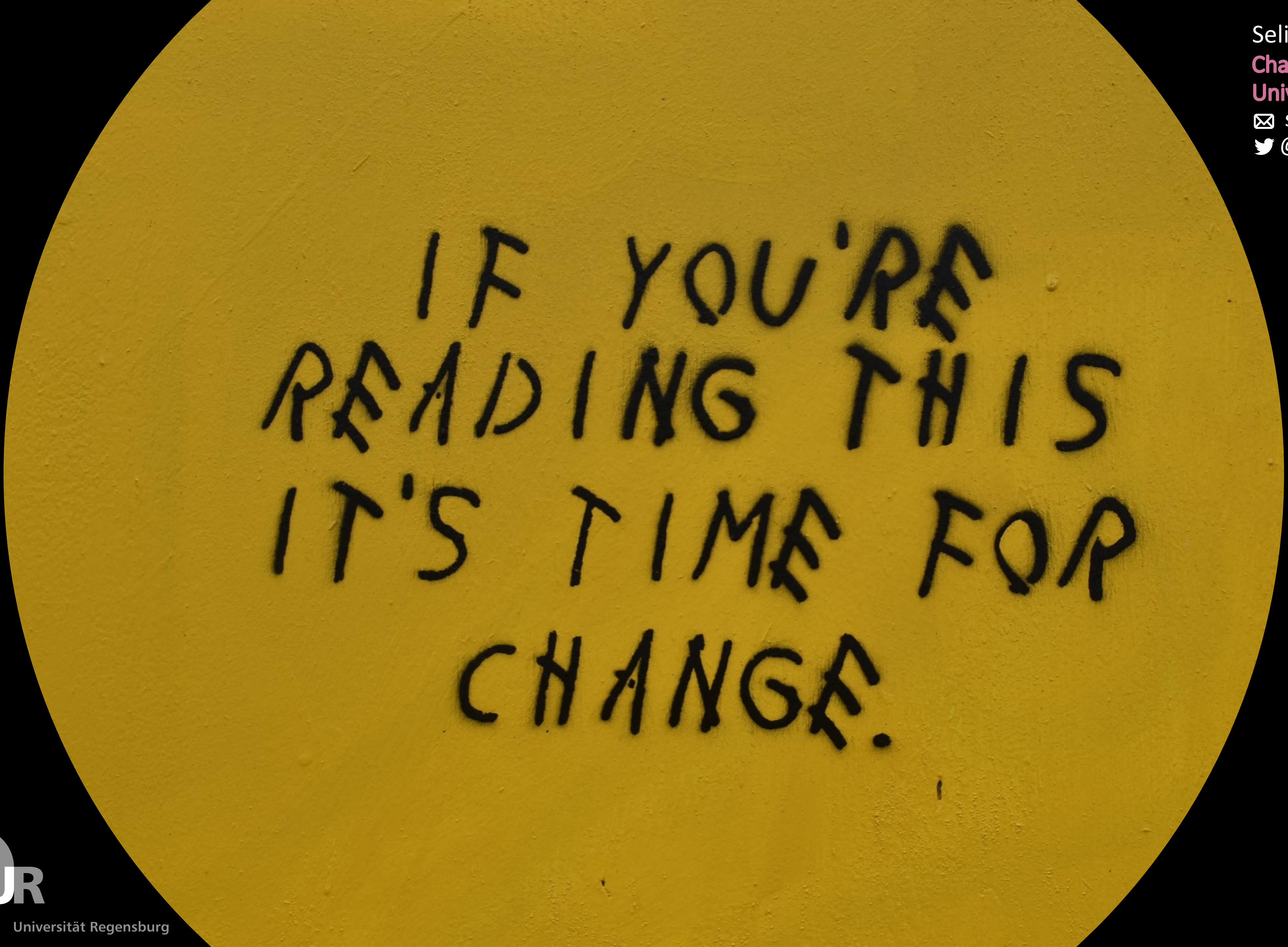
Illnesses resulting from poor health decisions have become leading cause of death

(Keeney, 2008)



Universität Regensburg

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey



Universität Regensburg

EUROPEAN LANGUAGE RESOURCES ASSOCIATION
ELRA

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Health behaviour changes are difficult to put into practice and sustain

(Kelly & Barker, 2016)

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Motivational Interviewing facilitates behaviour change

(Miller & Rollnick, 2002)

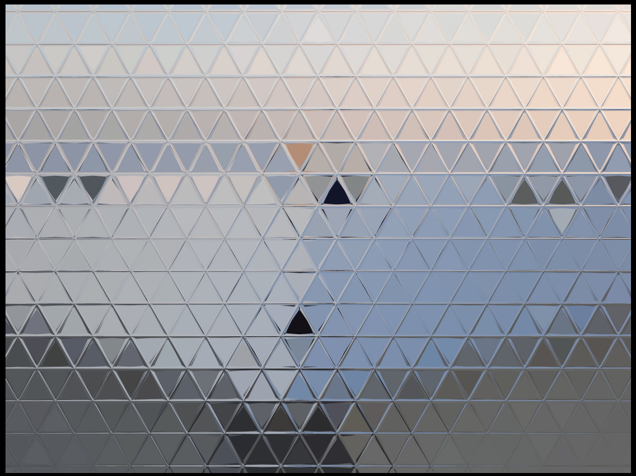
Automated delivery via a Conversational Agent has multiple potential benefits

(Lisetti et al, 2015)

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

A core aspect of MI is tailoring the conversation to the client and respond to their current state using

Open Questions



Reflections

Affirmations

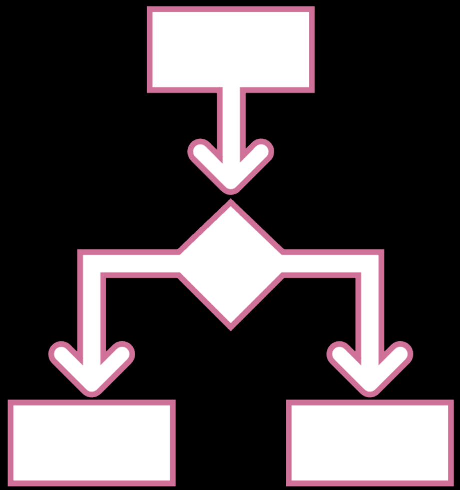
(Miller & Rollnick, 2002)

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Existing conversational agents rarely offer the flexibility needed for MI

- ○ Choice 1
- ○ Choice 2
- ○ Choice 3

Multiple choice input



Strict frameworks

Open Questions

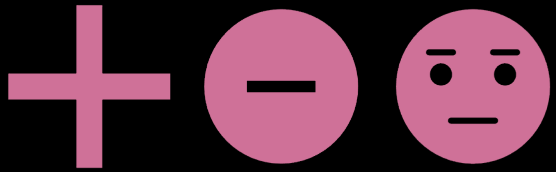
(i.e. Kocielnik et al, 2018; Bickmore et al, 2011, Nurmi et al. 2020)

Universität Regensburg

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Existing language resources focus mainly on therapist codes of spoken MI

only few resources include high level annotation of client behaviour



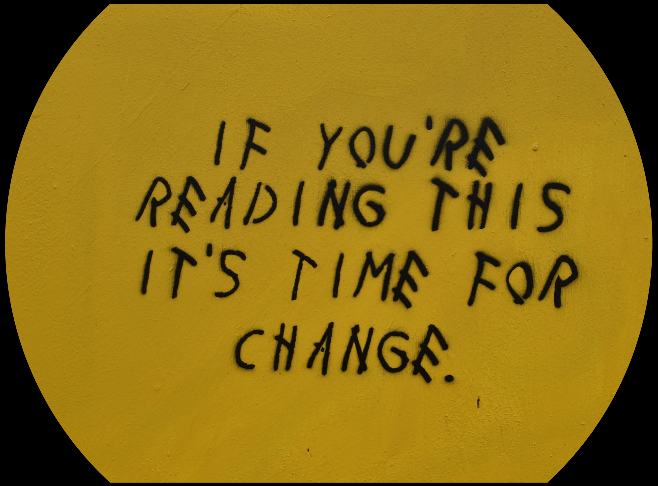
Universität Regensburg

(Pérez-Rosas, 2018; Tanana et al, 2016; Tavabi et al, 2020)

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

We are lacking suitable language resources of...

...written language...



...about behaviour change...

...with annotations of MI client behaviour codes

Relevant Client Codes in Motivational Interviewing

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

		Valence	
		Change Talk +	Sustain Talk -
Label	Reason	Rationale, basis, incentive, justification, or motive	
Label / Sublabel	desire	Desire or will	
Label / Sublabel	ability	Ability or degree of difficulty of the change	
Label / Sublabel	need	Necessity or need	
Label	Commitment	Agreement, intention, or obligation regarding future behaviour	
Label	Taking Steps	Specific steps that have been taken in the recent past	

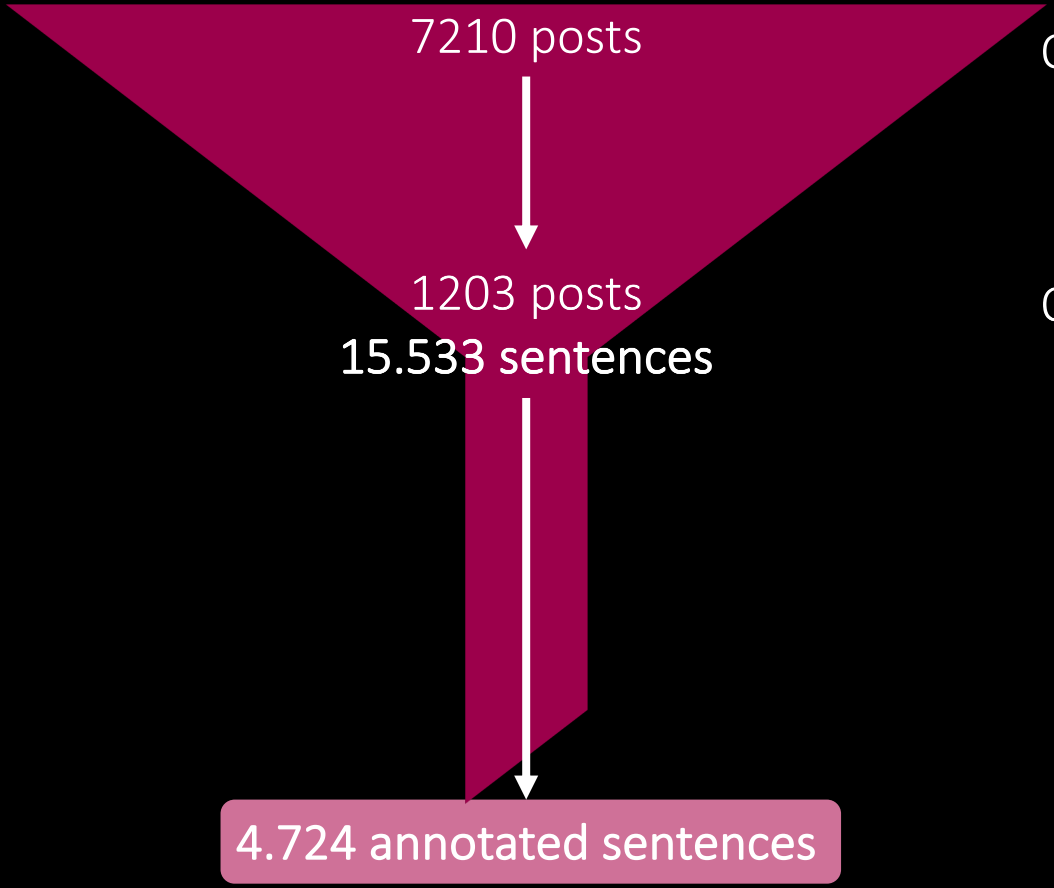
Follow/Neutral	Unrelated to speaker's current behaviour change
Other	Related to behaviour change but not attributable to other labels

(adapted from Miller et al, 2008)

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Data Collection

Data from two subforums of Germany‘s largest weight loss forum adipositas24.de



On Post level:

- Contains Change and/or Sustain Talk

On Sentence level:

- Client Codes
- Follow/Neutral (N=9643)
  - Often offering support/Information to others
- Other (N=828)
- Combination of Client Codes (N=321)

Universität Regensburg

EUROPEAN LANGUAGE RESOURCES ASSOCIATION ELRA

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Code Distribution
in Remaining Forum Data

		Valence	
		Change Talk +	Sustain Talk -
Label	Reason	28.3%	16%
Sublabel	desire	5%	0.9%
	ability	2.8%	7.4%
	need	3.8%	0.2%
Label	Commitment	9.2%	0.4%
	Taking Steps	20.1%	5.5%

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Data Analysis: Inter-Rater Reliability

Level	*Cohen's κ*
Valence	**0.755**
Label	**0.58**
R	0.621
TS	0.491
C	0.625
Sublabel	**0.654**
$R_{\text{no sublabel}}$	0.579
Ra	0.681
Rd	0.662
Rn	0.768

Inter-rater reliability scores are comparable to other research in the field

(Pérez-Rosas et al, 2016; Tanana et al, 2016; Hershberger et al, 2021)

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Data Analysis: Sentiment

- Classified randomly sampled sentences with pretrained German bert model for sentiment analysis (Guhr et al, 2020) and compared with valence annotations
- *Chi-Square* χ (2, N=1000) = 51.21, p < 0.0001); F1 = 27%

	negative	neutral	positive
-	217	81	23
+	298	265	116

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Data Analysis: Keywords

TS	C	R	$R_{\text{no sublabel}}$	Rn	Ra	Rd	+	-
have	will	is	have	must	can	want to	do	not
eaten	try	am	was	have to	hard	would like	hope	hard
eat	tomorrow	kg	am	important	manage	hope	now	problem
was	sometime	are		need	not	I	will	unfortunately
yesterday	first	fear		take care	manage	gladly	like	find
make	today	feeling		change	difficult	like	kilos	is
started	continue	yourself		work	find	wish	kg	nothing
changed	committed	satisfied		do	it	cake	goal	believe
have	go			find	know		finally	
day	next				doable			

Universität Regensburg

P < 0.05, Bonferroni corrected

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Machine Learning: Random Split

- Three datasets: Valence, Label, Sublabel
- Stratified random 80/20 Train-Test-Split
- Undersampling to the size of second largest class
- Finetuned GermanBERT to each dataset using 10-Fold-Cross-Validation
  - Three epochs, learning rate 5e-05

	Cross-Validation		Test Set		
	F1	**Std**	**Precision**	**Recall**	**F1**
Valence	73.97	2.63	70.42	73.31	70.87
Labels	74.16	3.22	79.64	74.87	76.96
Sublabels	79.49	2.69	66.20	81.89	71.53

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Machine Learning: Split by user activity level

- 65 most active users created 80% of dataset
- Used these 80% for training and texts by remaining 234 users for testing
- Do user specific utterances/conversational styles influence classification performance?

	Cross-Validation		Test Set		
	F1	Std	Precision	Recall	F1
Valence	75.11	2.24	72.39	74.76	72.86
Labels	76.31	3.78	71.38	73.71	72.46
Sublabels	79.43	2.6	62.84	74.76	66.69

Universität Regensburg

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Limitations

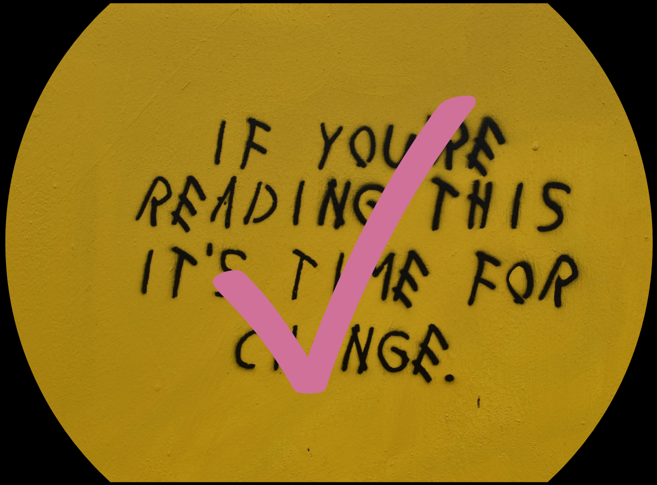
- Potential data bias towards people who are highly motivated to change
- Users were often seen to give advice and offer emotional support to others → Annotating therapist codes might yield further insight on when people share information
- Utterances containing multiple labels or annotated with Other | Follow/Neutral were not included in experiments and analysis

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Conclusion

https://github.com/SelinaMeyer/GLoHBCD

…written language…



…about behaviour change…

…with annotations of MI client behaviour codes

Future work will look into context and domain independent applications of classifiers trained on this data

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Related Work

- Keeney, R. L. (2008). Personal Decisions are the Leading Cause of Death. *Operations Research*, 56(6):1335–1347.
- Kelly, M. P. and Barker, M. (2016). Why is Chang- ing Health-Related Behaviour so Difficult? *Public Health*, 136:109–116.
- Miller, W. R. and Rollnick, S. (2002). *Motivational Interviewing, Second Edition: Preparing People for Change*. Applications of Motivational Interviewing Series. Guilford Publications.
- Lisetti, C., Amini, R., and Yasavur, U. (2015). Now all Together: Overview of Virtual Health Assis- tants Emulating Face-to-Face Health Interview Ex- perience. *KI-Künstliche Intelligenz*, 29(2):161–172.
- Kocielnik, R., Xiao, L., Avrahami, D., and Hsieh, G. (2018). Reflection Companion: a Conversational System for Engaging Users in Reflection on Physical Activity. *Proceedings of the ACM on Interac- tive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):1–26.
- Bickmore, T. W., Schulman, D., and Sidner, C. L. (2011). A Reusable Framework for Health Coun- seling Dialogue Systems Based on a Behavioral Medicine Ontology. *Journal of Biomedical Infor- matics*, 44(2):183–197.
- Nurmi, J., Knittle, K., Ginchev, T., Khattak, F., Helf, C., Zwickl, P., Castellano-Tejedor, C., Lusilla- Palacios, P., Costa-Requena, J., Ravaja, N., et al. (2020). Engaging Users in the Behavior Change Process With Digitalized Motivational Interview- ing and Gamification: Development and Feasibil- ity Testing of the Precious App. *JMIR mHealth and uHealth*, 8(1):e12884.

Universität Regensburg

Selina Meyer, David Elsweiler
Chair of Information Science
University of Regensburg
selina.meyer@ur.de
@selina_mey

Related Work

- Pérez-Rosas, V., Sun, X., Li, C., Wang, Y., Resnicow, K., and Mihalcea, R. (2018). Analyzing the Qual- ity of Counseling Conversations: the Tell-Tale Signs of High-Quality Counseling. In *Proceedings of the Eleventh International Conference on Language Re- sources and Evaluation (LREC 2018)*, pages 3742– 3748.
- Tanana, M., Hallgren, K. A., Imel, Z. E., Atkins, D. C., and Srikumar, V. (2016). A Comparison of Natural Language Processing Methods for Automated Cod- ing of Motivational Interviewing. *Journal of Sub- stance Abuse Treatment*, 65:43–50.
- Tavabi, L., Stefanov, K., Zhang, L., Borsari, B., Wool- ley, J. D., Scherer, S., and Soleymani, M. (2020). Multimodal Automatic Coding of Client Behavior in Motivational Interviewing. In *Proceedings of the 2020 International Conference on Multimodal Inter- action (ICMI '20)*, pages 406–413.
- Pérez-Rosas, V., Mihalcea, R., Resnicow, K., Singh, S., and An, L. (2016). Building a Motivational Interviewing Dataset. In *Proceedings of the Third Workshop on Computational Linguistics and Clini- cal Psychology*, pages 42–51.
- Hershberger, P. J., Pei, Y., Bricker, D. A., Crawford, T. N., Shivakumar, A., Vasoya, M., Medaramitta, R., Rechtin, M., Bositty, A., and Wilson, J. F. (2021). Advancing Motivational Interviewing Training with Artificial Intelligence: ReadMI. *Advances in Medi- cal Education and Practice*, 12:613.
- Guhr, O., Schumann, A.-K., Bahrmann, F., and Böhme, H. J. (2020). Training a Broad-Coverage German Sentiment Classification Model for Dialog Systems. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, pages 1627– 1632.

Universität Regensburg