Heliyon 10 (2024) e23825

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Understanding reminiscence and its negative functions in the everyday conversations of young adults: A machine learning approach

Andrea Ferrario [a,b,*], Burcu Demiray [c,d]

[a] ETH Zurich, Zurich, Switzerland
[b] Mobiliar Lab for Analytics at ETH, Zurich, Switzerland
[c] Department of Psychology, University of Zurich, Zurich, Switzerland
[d] Healthy Longevity Center, University of Zurich, Zurich, Switzerland

ARTICLE INFO

Keywords:
Reminiscence
Reminiscence functions
Young adults
Machine learning
Natural language processing
Well-being
Mental health
Digital health interventions

ABSTRACT

Reminiscence is the act of recalling or telling others about relevant personal past experiences. It is an important activity for all individuals, young and old alike. In fact, reminiscence can serve different functions that can support or be detrimental to one's well-being. Although previous studies have extensively investigated older adults' recalling of autobiographical memories, the evidence for young adults remains scarce. Therefore, in this work, we analyze young adults' production of reminiscence and their functions with a naturalistic observation method. Furthermore, we demonstrate that natural language processing and machine learning can automatically detect reminiscence and its negative functions in young adults' everyday conversations. We interpret machine learning model results using Shapley explanations. Our results indicate that young adults reminisce in everyday life mostly to connect with others through conversation, to compensate for a lack of stimulation or to recall difficult past experiences. Moreover, our models improve existing benchmarks from the literature on the automated detection of older adults' reminiscence in everyday life. Finally, our results may support the development of digital health intervention programs that detect reminiscence and its functions in young adults to support their well-being.

1. Introduction

1.1. Reminiscence across the adult lifespan

Reminiscence is the natural activity of thinking or talking about personally meaningful events from one's past [1]. It is an important part of our daily lives that serves a variety of purposes, such as providing individuals with a sense of identity, the possibility to review their lives and accept their past, the vision to solve current issues and the chance to enhance relationships [2,3]. Through the activity of reminiscing, we recollect memories of our self in the past, therefore accessing information belonging to our "autobiographical memory" [1]. This activity is particularly important for older adults' health and well-being: in fact, psychologists emphasize

* Corresponding author. ETH Zurich, Zurich, Switzerland.
E-mail address: aferrario@ethz.ch (A. Ferrario).

https://doi.org/10.1016/j.heliyon.2023.e23825
Received 15 February 2023; Received in revised form 11 December 2023; Accepted 13 December 2023
Available online 20 December 2023

reminiscence as a central task of old age and as promoting healthy aging [4]. Hence, there is a rich literature on the characteristics, functions and outcomes of reminiscence in both healthy and clinical samples of older adults [4–6].

However, lifespan developmental scientists argue that reminiscence is a routine aspect of the human experience and a common activity across the whole adult life span [7]. Thus, it is important to study the nature and functions of young adults' reminiscence activities, as well. To this end, different studies have found that one's age is unrelated to their frequency of engaging in reminiscence activities [3,6,8–12]. That is, similar to older adults, young adults have the need and will to review their experiences and give a meaning to them [13,14]. In fact, young adults access their sets of autobiographical memories and recall their past experiences to understand their sense of self, review their life goals and relationships as well as respond to everyday events [7,15]. Young adults' reminiscing is essential for them to ensure self-continuity and secure their well-being [15], develop their identity and a coherent life story [16]. Therefore, why and how young adults use autobiographical memories in their everyday lives may shed light on the relation between their reminiscing and well-being.

1.2. Functions of reminiscence in young adults

Reminiscing serves different functions—that is the purposes of recalling autobiographical memories [17]. Webster was the first to introduce a taxonomy of reminiscence that includes eight distinct functions, namely, *boredom reduction, death preparation, identity, problem solving, conversation, intimacy maintenance, bitterness revival* and *teaching/informing others* [18]. Webster's work [19] provides a compact description of these functions:

"*Boredom Reduction* measures our propensity to reminisce when our environment is understimulating and we lack engagement in goal-directed activities. *Death Preparation* assesses the way we use our past when thoughts of our own mortality are salient and may contribute to a sense of closure and calmness. *Identity* measures how we use our past in an existential manner to discover, clarify, and crystallize important dimensions of our sense of who we are. *Problem-Solving* taps how we employ reminiscence as a constructive coping mechanism […] *Conversation* measures our natural inclination to invoke the past as a means of connecting or reconnecting with others in an informal way […] *Intimacy Maintenance* measures a process whereby cognitive and emotional representations of important persons in our lives are resurrected in lieu of the remembered person's physical presence. *Bitterness Revival* assesses the extent to which memories are used to affectively charge recalled episodes in which the reminiscer perceives themselves as having been unjustly treated […] Finally, *Teach/Inform* measures the ways in which we use reminiscence to relay to others important information about life (e.g., a moral lesson)" [[19] pp.140].

Webster's taxonomy results from an instrument, called the Reminiscence Function Scale (RFS), that was developed collecting the responses of hundreds of adults [18]. The reminiscence functions are key to investigate the relation between individuals' reminiscing and their well-being [20–22]. To do so, the eight functions are typically clustered into three distinct groups: self-positive (*identity*, *problem solving*, *death preparation*), self-negative (*bitterness revival*, *boredom reduction*, *intimacy maintenance*) and prosocial (*conversation*, *teaching/informing others*) functions [21,23]. In particular, self-positive functions reaffirm previous self-understanding to develop self-awareness [22]. Importantly, they exhibit a positive association with physical health and psychological well-being [21,22]. Self-negative functions encode a continual dwelling on past experiences [22]. Notably, they show a negative association with physical and mental health [21,22]. Finally, prosocial functions affect physical and mental health, optimizing opportunities to experience positive emotions in social situations [22]. However, these results focus on older adults. They should feed into and encourage the investigation of young adults' reminiscence and how the relation between the functions of reminiscence and their well-being develops over time.

1.3. Into the wild: detecting reminiscence and its functions in everyday life

The literature on reminiscence and its functions is dominated by the retrospective self-report methodology. This methodology is used to elicit the recollection of autobiographical memories and their functions. To do so, experimenters typically ask study participants to complete Webster's RFS, indicating the degree to which they reminisce for each function of the provided taxonomy [20–22]. For example, in the clinical context of "life review" [2], which is a structured type of reminiscence in which one accesses memories to resolve conflicts and give meaning to current experiences [1], participants' autobiographical memories are elicited by therapists during ad-hoc sessions. A similar procedure is followed in more recent reminiscence-based interventions for young adults [13].

A first attempt to study reminiscence in real life is represented by Pasupathi and Carthensen's study that uses experience-sampling to characterize emotional experiences during mutual reminiscing [10]. However, more recently, different authors have started examining individuals' everyday life activities—including reminiscence—with naturalistic observation methods. One such method is the Electronically Activated Recorder (EAR) [24]. The EAR is a smartphone app that allows unobtrusively recording random snippets of participants' everyday conversations over time. It enables frequent, passive and privacy-preserving sampling of participants' language use in their natural environments via ambient recording [24–26]. Moreover, it allows preserving a high degree of naturalism, while collecting participants' acoustic logs throughout the days, resembling ethnographic methods [27]. At the time of writing, only one study had used the EAR to specifically investigate everyday life reminiscence activities—including their functions—and only with adults above age 62 [5]. By transcribing older adults' utterances verbatim from the conversations collected with the EAR, the authors were able to quantitatively characterize the production of older adults' reminiscence in everyday life and identify three reminiscence functions, as well as relate them to self-reported levels of life satisfaction [5].

Furthermore, research has recently started focusing on the possibility of automatically detecting cognitive activities (e.g., reminiscence), behaviors and environments using natural language processing (NLP) and machine learning (ML) on the transcripts of older adults' utterances from their everyday conversations that are collected with the EAR [28–31]. These studies show that it is possible to use ML and NLP to provide insights on the syntactic structure of the transcripts and its correlation with the manually-coded outcomes to be predicted, e.g., whether a given transcript of an older adult's utterance in a recorded conversation is a case of reminiscence. Further, their results provide evidence supporting the design of reminiscence-based digital health interventions to support older adults' well-being in their everyday life [28].

More recently, research also investigated language use in young adults' everyday conversations collected with the EAR [32,33]. However, authors did not consider young adults' reminiscence and its functions, focusing on the computation of measures of vocabulary richness and grammatical complexity, as well as lexical statistics on young adults' utterances transcribed from their conversations instead [32,33].

1.4. Our contributions

In summary, despite these recent efforts, research on the detection of young adults' reminiscence and their functions in everyday life is still at an infant stage. In particular, (1) quantitatively characterizing the production of young adults' reminiscence and its functions in their everyday life conversations, (2) detecting them automatically with ML and NLP methods, and, ultimately, (3) providing actionable recommendations on how to use these methods to support young adults' well-being, e.g., in technology-mediated reminiscence interventions, still remain open avenues of research. In this work, we make the following contributions:

1. We are the first to provide a quantitative analysis of the occurrence of reminiscence and its functions in young adults' everyday life conversations using the corpus of transcriptions of their utterances collected with the EAR technology in [34]. To do so, we aimed to answer the following questions:
   I. How much of young adults' utterances are reminiscence?
   II. Which functions does reminiscence serve in young adults' daily conversations?
2. We are the first to use ML and NLP to detect reminiscence and its functions using 3264 transcripts of young adults' everyday life conversations in German in two novel experiments. To do so, we (1) use four different families of ML models and their combinations in voting classifiers, (2) combine different families of features generated with NLP methods, (3) implement extensive hyper-parameter tuning, and (4) use different methods to cope with class imbalance. Furthermore, we provide insights on which NLP-generated features are most predictive of reminiscence using Shapley values [35], promoting the use of ML interpretability methods in social sciences.
3. We provide actionable recommendations on how to use our methodology to design digital health interventions that foster young adults' well-being by promoting self-reflection on their daily reminiscence activities automatically detected—together with their functions—with ML and NLP. In particular, we also show how the use of value sensitive design can address young adults' concerns and wishes (e.g., privacy) for digital health technology aiming to support their well-being. This, in turn, supports the adherence to the intervention program and its effectiveness.

2. Methods

2.1. Data collection and preparation

Data used in the current study have been originally generated in a study on "conversations and activities in everyday life" at the University of Zurich [5,34]. In what follows, we describe the data collection procedures in some detail and we refer to [34] for all details. We also note that these data have been analyzed by Luo and colleagues [32] with computational linguistic methods to examine the effects of context, namely familiarity with interlocutors, on young and older adults' language production in everyday life. However, Luo and colleagues did not examine reminiscence and its functions.

The original sample of young adults (age: 18–30) were recruited through the participant pool of the local department, snowball sampling and by different forms of advertising (i.e., flyers, online advertisement and on a local newspaper). Participants were mostly university students, who could choose between 50 Swiss Francs and research credits as compensation. All participants spoke Swiss German—i.e., an Alemannic dialect spoken in the German-speaking part of Switzerland—in everyday life.

The study included three sessions: a laboratory introduction session, a four-day observation period with the EAR, and a laboratory feedback session. First, participants were invited to the laboratory for an introduction to the study and were provided with an iPhone 4S with the EAR application (version 2.3.0) installed. For the observation period with the EAR, participants were asked to either clip the iPhone to their waistline or carry it in their pocket. They carried the iPhone with them for four consecutive days (one weekend and two weekdays, counterbalanced). The EAR was programmed to record 30-second-long audio files at random times, 72 times per day. This led to 288 audio files (72 $\times$ 4 days) and a total of 144 min (288 files x 30 s) per participant. The EAR was set to stay active for 18 h per day (blackout period between midnight and 6AM). Finally, participants were invited back to the laboratory for a feedback session, where they returned the iPhone. All study procedures were approved by the Ethics Research Institute of the Department of Philosophy at the University of Zurich. All participants gave written informed consent in accordance with the Declaration of Helsinki.

To protect the privacy of participants and their conversation partners, researchers followed the established guidelines for passive ambient audio sampling [27,36]. First, they limited the recording to a very small sample of the day (<5%). Second, the short recordings (30 s) ensured that minimal personal information was captured beyond what was necessary for reliable coding. Third, participants could review their data and delete any audio files they did not wish to share. Finally, they placed a "warning sign" on the iPhone to alert conversation partners of the possibility of being recorded (i.e., passive consent).

2.1.1. Data collection: transcribing and coding audio files

Demiray et al. [34] trained a team of research assistants to listen to each audio file, identify the participant's voice, and transcribe verbatim only the utterances of the participant (i.e., speech from non-participants was not transcribed to preserve their privacy). Assistants coded each file for (1) whether the participant was talking or not, (2) if talking, whether the participant was reminiscing or not, and (3) if reminiscing, which function(s) of reminiscence were present. All coding categories were dichotomous (0 vs. 1) indicating absence or presence of a behavior. Reminiscence referred to talking in detail about personally experienced past events that were meaningful to the participant: These could be specific events that happened at a particular place and time, repeated events (e.g., "I used to see him every weekend"), extended events (e.g., "our vacation in Italy was …"), and long periods of life (e.g., "When I was a university student …") [37]. Reminiscence functions were coded following the works of Webster et al. and Cappeliez et al. [18,23]. The same coding is also presented in Demiray et al.'s study [5]. The coded functions are *boredom reduction*, *death preparation*, *identity*, *problem solving*, *conversation*, *intimacy maintenance*, *bitterness revival* and *teaching/informing others*. They are not mutually exclusive, as noted in Demiray et al.'s study [5]. Two independent coders double-coded reminiscence and its functions. Inter-rater reliability for reminiscence was 87% and for reminiscence functions was between 87% and 100%. All sound files that showed a disagreement between the two coders were re-coded through discussion. As a result, the dataset in this work comprises 3264 coded transcripts of conversations from 66 young adults.

2.2. Natural language processing of transcripts

We used NLP to generate the following features from the coded transcripts: bag-of-words (BOW), part-of-speech (POS) tags and pre-trained word embeddings (EMB). The use of these features has been inspired by recent work on the detection of reminiscence, cognitive ability, social behavior and environments from older adults' conversations in everyday settings [28–31].

BOW generate transcript features by extracting all words in the corpus of transcripts and counting the word occurrences for all words in each transcript. POS tagging assigns a tag, called "POS tag", to each word extracted from a corpus of textual data [38,39]. The POS tags provide a description of the word role in the text, identifying whether it is a noun, an adjective, and so on. In this study, we used the tags provided by the POS-tagger in the Python library spacy. The tagger comprises 17 distinct POS tags, which we collect in the Appendix (Table A1). Similarly to BOW, the counts of unique POS tags are computed per each transcript. The counts of words and POS tags in each transcript are then normalized via "term frequency-inverse document frequency" (tf-idf) normalization [39] using the TfIdfVectorizer() function in the Python library sklearn.

Word embeddings (EMB) are numerical representations of words in a low-dimensional numerical vector with n components. The algorithms computing embeddings produce vector representations of words that occur in similar contexts close to each other, following the distributional hypothesis about languages and words [40]. The number of components n is fixed by the algorithm. In this work, we used the pre-trained word embeddings for German provided by spacy, selecting the model "de_core_news_sm", which computes the word embedding in an $n = 300$-dimensional space. Finally, the embedding of each transcript is given by the averaging the components of the embeddings of all its words [41].

2.3. Machine learning setting

We provide information on the ML setting by describing the (1) ML runs, (2) strategies against class imbalance, (3) cross-validation routine with recursive feature elimination, (4) hyperparameters in the cross-validation, (5) ML models and the (6) use of SHAP values to interpret ML model results. We will use this setting in two experiments involving the binary classification of the transcripts of the young adults' everyday conversations (as discussed in Section 2.4 and 2.5).

2.3.1. Machine learning runs

We considered seven different runs of ML modeling corresponding to different combinations of BOW, POS and EMB features. We collect them in Table 1.

2.3.2. Strategies against class imbalance

We used two strategies against class imbalance in the corpus of transcripts. The choice of the strategies is inspired by recent results on the detection of reminiscence and cognitive ability in older adults' conversations [28,29].

Table 1
All runs considered in this study.

Run	Feature combination
R0	BOW
R1	POS
R2	EMB
R3	BOW, POS
R4	BOW, EMB
R5	POS, EMB
R6	BOW, POS, EMB

In the first strategy, we performed class weighting (CW). This strategy reweighs data points during ML model training according to the class distribution in the training dataset. The reweighting results in the penalization of the cost of misclassifying data points from the minority class [42].

In the second strategy, we used data augmentation with the Synthetic Minority Oversampling Technique (SMOTE) algorithm. SMOTE allows generating synthetic data points from any numerical representation (i.e., BOW, POS, and EMB) of a transcript in the minority class by searching for its k-nearest neighbors ($k = 5$ is the default value) [43]. In particular, SMOTE has been used to detect reminiscence in the transcripts of older adults' everyday conversations with ML [28].

2.3.3. Cross-validation with recursive feature elimination

In each ML pipeline, due to the limited number of available data points and the high number of generated NLP features, we performed cross-validation with feature selection. To do so, we used the recursive feature elimination (RFE) algorithm [44] inside the cross-validation routine. RFE allows us selecting models with high performance but trained on a limited number of features. This becomes particularly relevant in all runs with BOW and EMB. The best model emerging from the cross-validation routines for each run (see Table 1) is selected using the F1 score, i.e., the harmonic mean of precision and recall [45]. In addition, we also computed the area under the Receiving Operator Curve (AUROC), precision and recall of the best models. With this ML pipeline, considering the SMOTE strategy, at each (stratified) partition of data into a training and validation fold, the ML models are trained on the balanced training data and their performance measures are evaluated on the imbalanced validation fold.

2.3.4. Hyperparameters in the cross-validation routine

Table 2 summarizes all hyperparameters tuned in the cross-validation routine that is considered in this work. We show the grids for all hyperparameters in the Appendix (Table A2).

2.3.5. Machine learning models

We considered four families of ML models: (1) eXtreme Gradient Boosting (XGB), (2) Light Gradient Boosting (LIGHT) [45–48], (3) support vector machines (SVM) and (4) random forests (RF). XGB and SVM have already been used for the detection of reminiscence, social behavior and environment from transcripts of daily conversations of older adults [28,30,31]. LIGHT is an efficient implementation of gradient boosting that delivered good results across different use cases [48]. In particular, together with RF it has been used to predict working memory in healthy older adults using real-life language and social context information from the transcripts of conversations in everyday life settings [29].

Finally, we considered voting classifiers by aggregating the best models of the different ML runs in a weighted ensemble [49]. In a binary classification problem, given a data point, each model of the voting ensemble computes two empirical probabilities. Then, for the given data point, the voting classifier returns the class whose empirical probability solves the optimization problem

$$\text{argmax}_j \sum_{i=1}^{N} \omega_i \, p_{ij},$$

where ω_i denotes the weight of the i-th model of the ensemble, p_{ij} the probability of the j-th class computed by the i-th model of the ensemble and $j = 1,2$. We trained voting classifiers considering the nine ensemble that comprise:

1. The four best-performing models for the CW and SMOTE strategies (separately),
2. The top three and two best-performing models for the CW and SMOTE strategies (separately),
3. The top four, three and two best-performing models across the CW and SMOTE strategies.

Note that the models participating to any of the nine ensembles may be trained on different sets of features. We searched for best performing voting models by tuning the weights ω_i of the models in the nine ensembles via grid search and selecting the model with highest F1 score on test data. In Appendix (Table A2), we show the grid for the weights used in the voting classifiers.

Table 2
Summary of all hyperparameters tuned in the cross-validation routine.

Pipeline element	Hyperparameters
SMOTE	number of neighbors
BOW features	n-grams
	stopwords
POS features	n-grams
RFE algorithm	number of features to select
	number of features to reduce at each step
ML models (XGB, LIGHT)	number of trees
	maximum tree depth
	learning rate
ML models (RF)	number of trees
	maximum tree depth

2.3.6. *Interpreting ML models with SHAP values*

Finally, we interpreted the results of the best ML models emerging from the cross-validation routines using Shapley Additive exPlanation (SHAP) values [35]. To the best of our knowledge, the use of SHAP values to interpret ML model results is still a novelty in social sciences research. However, examples of ML applications using SHAP values are found in medical practice and public health [50, 51], financial services [52], and affective computing [53] among others.

SHAP values are scores that fulfill a set of desirable properties, i.e., local accuracy, missingness and consistency [35], as opposed to other widely-used interpretability methods, such as Local Interpretable Model-agnostic Explanations (LIME) [54] and tree-based feature importance scores. They generalize the Shapley values from game theory [55] to the context of ML model explanations. A ML model becomes the payoff function of a collaborative game whose players are all possible sets of features used by the model to compute predictions. To do so, the SHAP algorithm computes the contribution of each feature, namely its SHAP value, to the ML model prediction at each data point. These predictions are explored individually, or aggregated to provide a global overview of the feature importance of the ML model at hand. In this study, we used two SHAP plots to interpret the ML models emerging from the cross-validation routines before voting: the SHAP global and beeswarm plots. The global SHAP plot provides an overview of the most important features by computing their mean absolute SHAP value on the dataset of transcripts. The beeswarm plot shows the distribution of SHAP values per each data point and ML model features instead.

We have run SHAP analyses on ML models *before* voting, as SHAP methods cannot be directly implemented in the case of voting classifiers whose ensembles comprise models trained on different families of features, which is the case for the models in this work.

2.4. *Experiment I: detecting reminiscence from the transcripts of everyday conversations*

In this experiment, we implemented the ML setting to detect reminiscence utterances in the corpus of 3264 transcripts of young adults' everyday conversations (see Fig. 1). For both strategies against class imbalance, we performed a stratified partition of the corpus of transcripts in training and test data with 80:20 ratio. Then, we implemented a $k = 10$ cross-validation routine with RFE on the training data to select the best model, computing its performance on test data. As the model performance is computed only on one (test) dataset, we also computed the 95% bootstrapped confidence intervals of the F1 score of the best model using 1000 iterations [56]. We used different SHAP value plots to interpret the feature importance of the best model across the two strategies. Finally, we combined the best models of the two strategies into the nine distinct voting classifiers discussed in Section 2.3.5.

2.5. *Experiment II: detecting negative reminiscence events in the corpus of reminiscence transcripts*

In this experiment, we implemented the ML setting to detect negative reminiscence events in the sub-corpus of all reminiscence utterances. To do so, we classified a reminiscence transcript as "negative", if it has been coded as "self-negative" function (i.e., *bitterness revival*, *boredom reduction*, *intimacy maintenance*) or *death preparation*. However, differently from Experiment I, due to the low number of reminiscence samples, we performed no split into training and test data, implementing a $k = 5$ cross-validation routine with RFE on the whole sub-corpus, as shown in Fig. 1, for both strategies against class imbalance. As a result, we computed the mean and standard deviation of the model performance metrics on the $k = 5$ validation folds, for the best model of each strategy. Then, we used different SHAP value plots to interpret the feature importance of the best model and we combined the best models of the two strategies into the nine distinct voting classifiers discussed in Section 2.3.5.

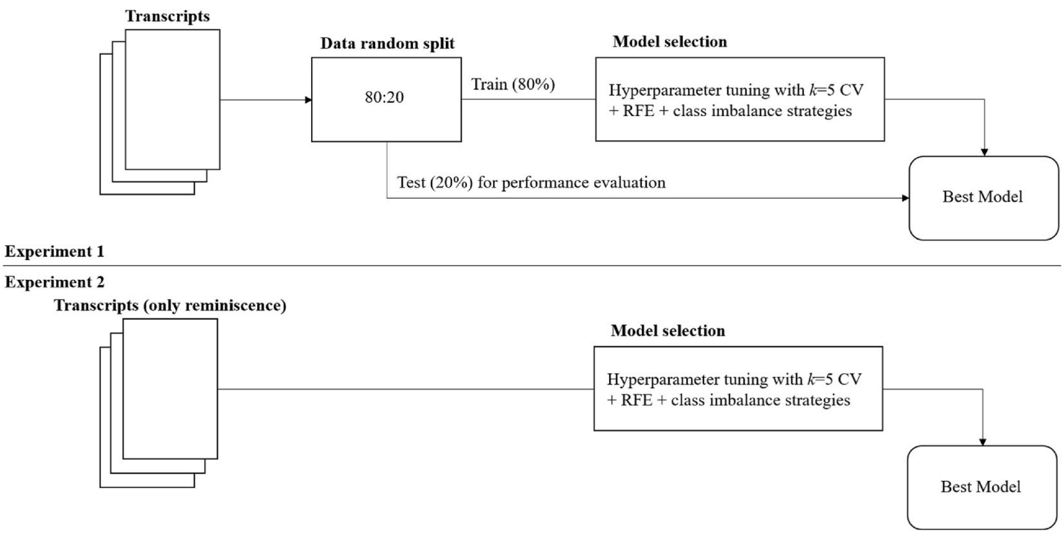

Fig. 1. The ML pipelines for the two experiments in this work.

3. Results

3.1. Descriptive analysis of the dataset

The dataset comprises of 3264 transcripts from 66 participants. On average, we collected 49 (SD = 30, range = 1–127) transcripts from each participant. There were a total of 219 transcripts with reminiscence (6.7% of all transcripts). Out of 66 participants, 52 engaged in reminiscence at least once. On average, participants uttered 4 (SD = 3, range = 1–13) cases of reminiscence. In Table 3, we show the distribution of the reminiscence functions in the dataset of all transcripts. Most reminiscence cases served the *conversation* function, followed by the negative functions *boredom reduction* and *bitterness revival*. The remaining functions are either rare (i.e., under the 5% quota) or absent. This latter is the case for *death preparation* and *intimacy maintenance* functions.

Clearly, the coded functions of reminiscence are not mutually exclusive. In fact, as shown in Table 4, 54 reminiscence cases (25%) are coded with two functions, one of which is always *conversation*.

All transcripts coded as *bitterness revival* or *problem solving* are also coded as *conversation*. However, not all *boredom reduction*, *teaching/informing others* and *identity* reminiscence cases are coded as *conversation*, as well. Given the above overview on reminiscence and its functions, in Table 5, we show a few transcripts, providing an example for each reminiscence function.

3.2. Experiment I

3.2.1. Summary of the best ML model runs

In Table 6 and 7, we collect the best models resulting from the ML pipeline used in Experiment I, as described in Section 2.4. Only 6.7% of transcripts are coded as reminiscence.

The best model for the CW strategy is a LIGHT ensemble of 50 trees with depth equal to ten, and learning rate equal to 0.1. The RFE algorithm selects only 30 features, deleting 25% of the features at each step. The ensemble is trained using 1-gram BOW and POS-tags features, without the removal of stopwords. It reaches F1 = 0.552, CI = [0.449, 0.579], outperforming all other classifiers. The best model for the SMOTE strategy is an XGB ensemble of 100 trees of depth equal to five and learning rate equal to 0.1, instead. The RFE algorithm selects only 30 features from POS-tags and word embedding components, deleting 25% of the features at each step. The ensemble is trained using 1-gram POS-tags and word embeddings on SMOTE-augmented data with $k = 13$ nearest neighbors. This model reaches F1 = 0.545, CI = [0.450, 0.590] outperforming all other classifiers.

Finally, we note that combining the best models in a voting classifier further improves performance. In fact, as shown in Table 7, the voting classifier comprising the best XGB (SMOTE strategy, weight $\omega_1 = 0.3$) and LIGHT (CW strategy, weight $\omega_2 = 0.7$) models reaches F1 = 0.577, CI = [0.469, 0.611], i.e., an improvement equal to 4.5% with respect to the F1 of the best LIGHT model (CW strategy). In particular, this voting classifier improves precision with respect to all models and class imbalance strategies.

3.2.2. SHAP values analysis

We interpret the best LIGHT (CW strategy) model from Table 6 using SHAP values analysis.

The LIGHT best model is trained on 30 BOW and POS features. In Fig. 2, we show the SHAP global and beeswarm plots for its 20 most important features. The SHAP global plot indicates that the most important features are the POS-tag "AUX", i.e., auxiliary verbs,

Table 3
Distribution of the reminiscence functions in the dataset (219 reminiscence).

Type	Function	Counts (%)
Negative	boredom reduction	27 (12%)
	bitterness revival	15 (7%)
	intimacy maintenance	–
Positive	identity	9 (4%)
	problem solving	4 (2%)
	death preparation	–
Social	conversation	208 (95%)
	teaching/informing others	9 (4%)

Table 4
Distribution of reminiscence cases with multiple functions. The percentages are respect to the total number of reminiscence transcripts that are coded with the first function (see Table 3).

Function	Counts (%)
boredom reduction AND conversation	26 (96%)
bitterness revival AND conversation	15 (100%)
teaching/informing others AND conversation	8 (89%)
identity AND conversation	6 (67%)
problem solving AND conversation	4 (100%)
(intimacy maintenance AND conversation) OR (death preparation AND conversation)	–

Table 5
A few examples of reminiscence from the dataset of all transcripts. All examples are translated from German into English using Google Translator, followed by a review by the authors.

Type	Function	Example
Reminiscence	boredom reduction	"Yes, I just totally hate science fiction. Yes. Totally, completely, *mega*. Yes, last time, my roommate [asked me] 'are you coming to watch [The] Hobbit?' [I answered] 'What, I'm not coming to see [The] Hobbit'. No, well I have never watched it. I watched Harry Potter and …"
	bitterness revival	"[I] said 'easy, come train with us, then you'll have some distraction and stuff like that'. He was, you know, upset about you, he was almost more upset than I was about you. But it doesn't matter what your friends think. I don't always want to know what my friends think of me either. Probably not always the best either. Yes so. And it can't go on like this, you know that too."
	identity	"It's been a year now and I've also been completely promoted. And I just don't want to go there anymore."
	problem solving	"Hi. Uh, I have a question that is I have a MacBook Air and I had it in my bag and some water spilled in the bag. Then, […] I mean the MacBook, then I immediately turned it off and put in white rice just to suck off the water."
	conversation	"At some point I bought a shampoo at [REDACTED]. Somehow I have the feeling that it does not wash [my hair] properly. My hair smells exactly the same [as before]"
	teaching/informing others	"Yes, so what was that, a wasp or a bee? I do not know either, it stabbed me in there and then it gave [me] blood poisoning. It really swelled up there and one could not recognize the finger, it was just round. And back then we did not know that I was allergic, it was about five years ago. And then we said, "come on, let us wait a minute, it will go back anyway' and afterwards it really got worse and worse and afterwards …"
Not reminiscence		"Oh, I still need milk. Do you still have some? Good, yes."

Table 6
Experiment I: Performance of the best models for all strategies against class imbalance.

Model	Strategy	Features	AUROC	Precision	Recall	F1
XGB	CW	BOW	0.916	0.413	0.705	0.521
LIGHT		**BOW, POS**	**0.932**	**0.475**	**0.659**	**0.552**
SVM		BOW, POS	0.910	0.341	0.705	0.459
RF		POS, EMB	0.931	0.436	0.545	0.485
XGB	SMOTE	**POS, EMB**	**0.935**	**0.455**	**0.682**	**0.545**
LIGHT		BOW, POS	0.931	0.424	0.568	0.485
SVM		BOW, POS, EMB	0.922	0.368	0.727	0.489
RF		POS, EMB	0.940	0.435	0.614	0.509

Table 7
Experiment I: Performance of the best voting models. In brackets the weights of the classifiers in each ensemble.

Ensemble	Strategy	AUROC	Precision	Recall	F1
XGB (0.3) LIGHT (0.7)	CW	0.936	0.492	0.659	0.563
XGB (0.9) SVM (0.1)	SMOTE	0.942	0.462	0.682	0.550
CW LIGHT (0.7) SMOTE XGB (0.3)	CW and SMOTE	**0.938**	**0.528**	**0.636**	**0.577**

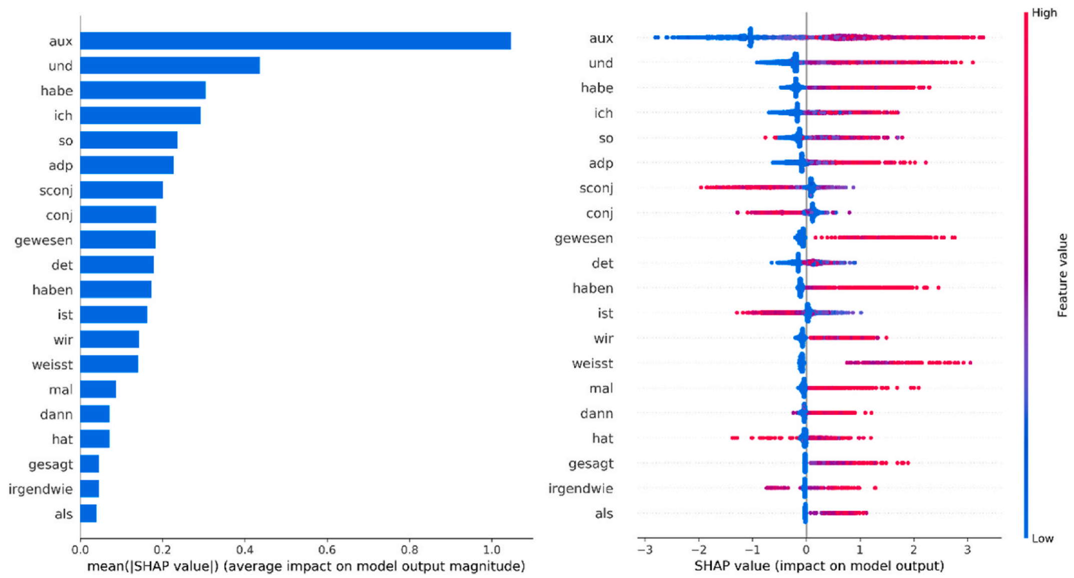

Fig. 2. SHAP global plot (left) and SHAP beeswarm plot (right) for the best LIGHT model (CW strategy). In the global plot, the importance is computed as the absolute value of the mean SHAP value of each feature. In the beeswarm plot, the horizontal axis represents the Shapley values. The dots are jittered vertically and colored according to the feature value.

and the tokens “und” (“and”), “habe” (”[I] have”) and “ich” (“I”). In the SHAP beeswarm plot we show the distribution of SHAP values per each transcript in the dataset and each of the 20 features of the SHAP global plot, instead. We note the presence of clusters of observations (depicted as blue vertical bars) for all features. These clusters of data correspond to all transcripts in which the corresponding feature has value equal to zero. By definition of tf-idf normalization of BOW and POS features, this is equivalent to state that in those transcripts the feature, i.e., the token originated via BOW or POS-tagging is missing. For example, 34% of all transcripts contain no auxiliary verb, i.e., “AUX” $= 0$. All these transcripts are not coded as reminiscence. In summary, the beeswarm plot indicates that for the top-six most important model features, their absence in a transcript leads to a negative corresponding SHAP values and, therefore, it is negatively correlated with its probability of being a reminiscence, by definition of SHAP values. Moreover, it shows that high values of the POS tags “AUX”, “ADP”, i.e., adpositions (prepositions and postpositions), and the tokens “und”, “habe” and “ich” are positively correlated with the probability of a transcript to be a case of reminiscence. By definition of the tf-idf normalization, the high feature values could be the result, for example, of a high frequency of these POS and tokens in a given transcript. On the contrary, high values of the POS-tags “CONJ”, i.e., conjunctions, “SCONJ”, i.e., subordinating conjunctions (“as long as”, “because”, “unless” etc.), and the token “ist” (”[he/she/it] is”) are negatively correlated with the probability of a transcript to be a reminiscence.

3.3. Experiment II

3.3.1. Summary of the best ML runs

In Table 8 and 9, we collect the best models resulting from the ML pipeline used in Experiment II, as described in Section 2.5. Only 12.2% of the 219 reminiscence cases are labelled as "negative".

Considering the CW strategy, the RF model trained on 1- and 2-gram BOW and EMB features outperforms all other classifiers, reaching F1 = 0.460 (0.082). The forest comprises 50 tree stumps and it is trained on ten features selected by the RFE algorithm by deleting 50% of all features at each step. Stopwords are removed. Considering the SMOTE strategy, the LIGHT model trained on EMB features and comprising an ensemble of ten tree stumps outperforms the other classifiers, reaching mean F1 = 0.486 (0.146). The RFE algorithm selects only five EMB features, deleting 25% of the features at each step. The model is trained on SMOTE-augmented data with $k = 11$ nearest neighbors.

Furthermore, using voting improves performance. In fact, as shown in Table 9, the voting classifier comprising the best (1) LIGHT (SMOTE strategy, weight $\omega_1 = 0.15$), (2) RF (strategy CW, weight $\omega_2 = 0.15$), (3) XGB (SMOTE strategy, weight $\omega_3 = 0.65$) and (4) SVM (CW strategy, weight $\omega_4 = 0.05$) models reaches mean F1 = 0.543 (0.077), i.e., +11.7% in mean F1 with respect to the best LIGHT model for the SMOTE strategy. In particular, this voting classifier shows an increase in mean precision equal to 21.3% with respect to the LIGHT model. However, as the best model (before voting) is trained on EMB features only, we do not perform a SHAP value-based analysis of its feature importance distribution. In fact, the $n = 300$ components of the pre-trained word embedding model do not hold a direct interpretation, differently from BOW tokens and POS tags. Similarly, a standard feature importance analysis of the best RF model for the CW strategy shows that all ten features selected by the RFE algorithm are EMB.

4. Discussion

4.1. Reminiscence production and its functions

Reminiscence is a process engaged in by adults of all ages [3,12]. Our study shows that the frequency of reminiscence in everyday conversations of young adults is equal to 6.7%. This result is comparable to the frequency of reminiscence in older adults (5% [5]) and suggests that reminiscence is an intrinsic developmental resource and a common activity for young adults, as well.

In terms of reminiscence functions, we found that young adults reminisced mostly for the *conversation* function (95%). This is not surprising, as our data come from actual conversations in everyday life. However, it also indicates that reminiscence is indeed a resource that people use to socialize and connect with others in everyday life. Next, the young adults in this work did not reminisce with the *death preparation* or *intimacy maintenance* functions. This indicates that death-related functions may be irrelevant for the developmental tasks in young adulthood or may not play a big role in the mental worlds of young individuals.

Furthermore, the reminiscence conversations of our participants rarely served the *problem solving* and *identity* functions. This result diverges from previous literature on these functions for young adults where, however, autobiographical memories were recalled with self-report methods [3,18,20]. Similarly, *teach/inform others* is served rarely; this is in line with previous results showing that this function increases with age, showing a plateau at age 40 [20].

Finally, the higher frequency of two self-negative functions (i.e., *boredom reduction* and *bitterness revival*) in the corpus of everyday conversations of our study is in line with previous work on young adults' reminiscing, albeit with different reminiscence elicitation methods [17,20]. Graham et al. (2020) used Webster's RFS and person-centered statistics in a study with 907 participants (mean age = 32.83, SD = 18.01, range = 17–88) to identify age- and function-related patterns of reminiscence [17]. They found three distinct reminiscing profiles. In particular, the "Young-Adult Self-Negative" profile is characterized by the lowest age (mean = 20.14, SD = 1.79) and the highest scores on *boredom reduction* and *bitterness revival* [17]. Therefore, our results are in line with the "Young-Adult Self-Negative" profile, in the sense that they show that young participants reminisced the most for *boredom reduction* and *bitterness revival*.

In summary, our results suggest that, considering a naturalistic observation method and everyday conversations, young adults may use reminiscence in a conversational setting to cope with present negative contexts (i.e., boredom) and to analyze or make meaning of negative past experiences (the revival of bitterness) more than for other functions. Future work could analyze differences in the functions of reminiscence between young and older adults when considering their conversations in everyday life, similarly to existing attempts in the literature that rely on self-report [20,21]. Furthermore, future psychological work could investigate why young adults tend to rely mostly on negative functions while reminiscing and how this is associated with their well-being.

Table 8
Experiment II: Mean performance of the best models on $k = 5$ validation folds, for both class imbalance strategies. In brackets the standard deviations.

Model	Strategy	Features	AUROC	Precision	Recall	F1
XGB	CW	BOW, POS, EMB	0.609 (0.060)	0.371 (0.084)	0.544 (0.102)	0.428 (0.055)
LIGHT		BOW, POS, EMB	0.640 (0.091)	0.430 (0.103)	0.450 (0.122)	0.429 (0.085)
SVM		BOW, POS, EMB	0.691 (0.056)	0.368 (0.077)	0.619 (0.087)	0.451 (0.044)
RF		**BOW, EMB**	**0.656 (0.051)**	**0.368 (0.073)**	**0.617 (0.100)**	**0.460 (0.082)**
XGB	SMOTE	EMB	0.651 (0.062)	0.339 (0.082)	0.712 (0.058)	0.454 (0.079)
LIGHT		**EMB**	**0.720 (0.100)**	**0.417 (0.128)**	**0.592 (0.189)**	**0.486 (0.146)**
SVM		EMB	0.677 (0.080)	0.318 (0.065)	0.639 (0.143)	0.423 (0.085)
RF		BOW, POS, EMB	0.649 (0.049)	0.335 (0.054)	0.569 (0.130)	0.421 (0.078)

Table 9
Experiment II: Mean performance of the best voting models on $k = 5$ validation folds. In brackets the standard deviations.

Ensemble	Strategy	AUROC	Precision	Recall	F1
RF (0.65), SVM (0.05), LIGHT (0.25), XGB (0.05)	CW	0.680 (0.047)	0.446 (0.152)	0.475 (0.141)	0.452 (0.134)
LIGHT (0.15) XGB (0.65) SVM (0.15) RF (0.05)	SMOTE	0.725 (0.086)	0.453 (0.103)	0.667 (0.115)	0.535 (0.100)
LIGHT SMOTE (0.15) **RF CW (0.15)** **XGB SMOTE (0.65)** **SVM CW (0.05)**	CW and SMOTE	**0.748 (0.085)**	**0.506 (0.101)**	**0.594 (0.059)**	**0.543 (0.077)**

4.2. *Experiment I*

Our results show that gradient boosting methods, i.e., LIGHT and XGB, outperform all other families of ML models. Moreover, aggregating the predictions of boosting methods in an ensemble further enhances their performance. This is due, in particular, to an increase in precision. Our results improve previous work on the detection of reminiscence from transcripts of everyday conversations, albeit of older adults, as no benchmark exists for the case of young adults. In fact, Ferrario et al.'s best SVM (CW strategy) achieves F1 = 0.480, although on BOW features, only [28]. This result is improved by Stoev et al. who achieved F1 = 0.519 using a RF model trained on BERT-augmented data and different families of NLP features, including EMB, on the same dataset [30]. Indicatively, our best (voting) model improves Stoev et al.'s results with an 11.2% increase in F1 score. Despite this improvement, and similarly to other studies on the detection of reminiscence, cognitive ability, behavior and environment using older adults' transcripts everyday conversations studies [28–31], our models are affected by rather low precision on test data. Arguably, this is due to the low number (i.e., 44) of reminiscence transcripts and the class imbalance in test data. In particular, the false positives that affect precision comprise long transcripts with sentences in different tenses, including the past.

The SHAP value analysis states the importance of auxiliary verbs, i.e., imperative, infinitive and perfect particle of "to be", "to have" and "to become", in relation to the probability of a transcript to be a case of reminiscence. This, together with the importance of the tokens "ich", "und" and past particles, such as "gewesen" ("been"), "gesagt" ("said"), suggests that ML models characterize young adults' reminiscence as a syntactically articulated utterance describing a past first person activity, coherently with the definition of autobiographical memory. Moreover, the importance of auxiliary verbs and German stopwords (e.g., "ich", "habe", and "gewesen") replicates previous results on the detection of reminiscence from older adults' everyday conversations [28]. These findings suggest that young and older adults' reminiscing in everyday conversations share syntactic similarities.

4.3. *Experiment II*

Our results show that it is possible to classify negative reminiscence functions in a corpus of reminiscence transcripts, although the limited number of data points and class imbalance affect modeling performance. In fact, the mean F1 score of all models is computed over the $k = 5$ (class imbalanced) validation folds, which contain, on average, only five reminiscence transcripts coded as "negative". Differently from Experiment I, RF models outperforms all boosting methods for the CW strategy. The interpretability of the RF model results is however hindered by the use of EMB features only. Similarly to Experiment I, ensembling further improves performance, due to an improvement in mean precision. In fact, the voting classifier comprising all four best models (see Table 9) achieves performance that is in line with the one of the best XGB (SMOTE strategy) model from Experiment I (see Table 6), although a direct comparison of the two models is not possible.

We note that the proposed methodology uses only the transcriptions of the utterances of the study participants and their reminiscence-related codes. Namely, it does not intercept the social contexts or behaviors that characterize conversations as opposed to, for example, methods, such as conversation analysis combined with the coding of social interaction [57,58]. In particular, the coding of social interaction allows quantifying the behavior observed with conversation analysis and testing relationships between coded interactions and other variables, such as sociodemographic ones, those collected by self-assessment questionnaires or selected outcomes [58]. Therefore, future research on the automated detection of young adults' reminiscence and its functions may consider the formal coding of the social behavior and interactions and use this information in ML modeling pipelines. Recently, Ferrario et al. used a similar approach, investigating whether sociodemographic information, language use and the "social context" of everyday conversations predict cognitive measures of healthy older adults, using ML and NLP methods [29]. In particular, they coded 19 variables to describe the social context of a conversation [29]. "Context" comprises the environment in which the conversation takes place, the type of conversation and the older adult's partner and, finally, the activities during which the conversation unfolds [29].

Finally, we note that experiment I and II results suggest that an "accuracy-interpretability" trade-off [59] exists when using ML models to detect reminiscence and its negative functions. In fact, although ensembling models in voting classifiers improved performance in both experiments, the impossibility to use SHAP value-based analyses of the voting predictions due to the fact that the models in the ensembles are trained on different sets of features, lowers the overall interpretability of the voting classifiers. However, there is no one-size-fits-all solution to this trade-off: depending to the application—for example, digital health interventions, see

Section 4.4.—researchers have to weigh the need for high model performance vs. the right of different stakeholders to understand the outputs of the models and act upon them.

4.4. Reminiscence-based digital health interventions to enhance young adults' well-being

The use of reminiscence in interventions and therapies for older adults is rather common, emphasizing the relation between positive functions of reminiscence and psychological well-being [60] and the association between self-negative functions and poorer mental health outcomes [21]. These interventions typically distinguish between (1) simple reminiscence, namely unstructured autobiographical storytelling in group or one-to-one format, (2) life-review and (3) life-review therapy [61].

Pinquart & Forstmeier's meta-analysis of 128 studies show that reminiscence interventions affect a broad range of outcomes (e.g., depressive symptoms, psychological well-being, ego-integrity and social integration) [60]. Interestingly, the positive immediate effects on different indicators of mental health and positive well-being are maintained also at follow-up [60]. However, as reported by the authors [60], the participants to the 128 interventions had mean age of 73.1 years (SD = 12.7; range = 18.8–85.7). As a result, although there is a rather large literature on the effect of reminiscence interventions on older adults, little work is focused on interventions specifically for young adults [13,62]. A recent, technology-mediated example is given by Hallford et al.'s work where the authors consider a reminiscence-based intervention that is delivered with an online, teleconference format and focuses on cognitive-reminiscence therapy and depressive symptoms [13]. This is an important research gap, as different authors challenge the exclusive use of reminiscence-based therapies with older adults and promote them with young adults, arguing that they are effective in younger populations as well [13,14,63]. In addition, the findings of our study specifically point to the self-negative functions observed with high frequency in young adults' conversations, raising questions such as "How might these negative functions affect young adults' well-being?" and "Could young adults be trained to reminisce in more positive or functional ways?". These reflections are especially important, as young adults typically engage in the recalling of events that could threaten how they view themselves and the world around them [15].

We believe that technology can play a pivotal role in delivering effective reminiscence-based therapies targeting young adults. Interestingly, different authors have discussed the benefits of technology in supporting rich and engaging reminiscence experiences of older adults [64,65]. A first attempt to use technology for reminiscence in young adults' everyday life is represented by tools, such as *Pensieve*, where users' social media content are used as triggers for reminiscing [66]. However, despite these initiatives, the promotion of technology-mediated reminiscence specifically for younger adults is still an open avenue of research. In general, technology can be leveraged to design digital health intervention programs that foster young adults' well-being by promoting self-reflection on their daily reminiscence activities and functions. In particular, these interventions could also be used to manage severe mental challenges, such as depression, and be integrated by the supervision of therapists to further improve their benefits [67]. As digital interventions are rather affordable and scalable, they can potentially be used by a large number of young adults.

In particular, the methods discussed in this work, namely the EAR in combination with ML models and NLP, could be used to design an example of such an intervention. In fact, reminiscence events and their functions could be detected and monitored (using the EAR and high-performance ML models, including those for the automated coding of conversations), analyzed over time (with time- and context-personalized feedback and "daily diaries" providing a statistical summary of daily reminiscence events and functions) and managed to promote well-being (with a focus on recalling and analyzing self-positive functions and reviewing the everyday life contexts and mental processes that characterize negative ones). In particular, monitoring statistics and user notifications, e.g., micro-interventions [68], could be delivered using a smartphone app. Designers could then measure the effectiveness of an intervention quantifying its effect over time on depression and quality of life indicators, such as meaning in life, positive well-being and social integration, similarly to existing reminiscence interventions [28,60]. Furthermore, the use of smartphone sensors data (e.g., Wifi, GPS, accelerometer and phone use) may provide valuable information to analyze the relation between reminiscence frequency and its functions and young adults' physical and social activities [69].

However, we are aware that the introduction of digital technology, especially for managing the effects of highly personal autobiographical memories on well-being, may give rise to various user concerns. These concerns may highlight potential threats posed by digital sensing technologies and ML methods, such as those affecting users' well-being, privacy, autonomy and identity [70]. For example, some young adults may be worried about the management of their conversation data, the lack of explanations of ML models used to detect reminiscence and trigger personalized notifications, the overall trustworthiness and efficacy of the digital intervention and trust in it [71].

Others may argue that a technology-intensive approach to self-reflection is a source of "technostress" and may prefer other types of interventions [72]. Although the EAR method has established protocols providing ethical safeguard measures and a low level of obtrusiveness [27], designers should take care of young adults' expectations, wishes and concerns on these digital health interventions more in-depth. To do so, they may consider a framework, such as value sensitive design (VSD), that provides a three-placed methodology accounting for values in the design process of a technology by promoting a (1) conceptual, (2) empirical and (3) technical investigation of system design [73]. Although VSD has been recently used for applications assessing quality of life of people with mental health problems [74], digital assistance for physiotherapeutic treatments [75] and digital health interventions to manage stress at the workplace [53], the use and testing of VSD methods for the design of digital interventions focused on autobiographical memories are still an open avenue of research.

4.5. Limitations

This study has several limitations. The dataset has a limited number of reminiscence cases (6.7% of transcripts), which was generated by a naturalistic observation study comprising only four days of data collection and 66 participants. Although the dataset in our study is larger than the one of similar studies on the detection of older adults' reminiscence [28,30], the rather limited sample size affects the performance of the ML models. In particular, this limitation does not allow the use of data-intensive ML methods, such as neural networks, to detect reminiscence and its functions. Therefore, future research may also consider the use of large language models to generate high-quality synthetic examples of transcribed young adults' autobiographical memories [76]. However, the use of larger sets of data makes the manual transcription and coding of conversations unfeasible. Therefore, (semi-) automated methods have to be considered to scale the proposed approach [31]. Further, although the EAR method allows recording snippets of conversations, data used in this study comprise only the transcripts of the utterances of the study participants and their reminiscence-related coding. The goal of this study was to assess the feasibility of detecting reminiscence and its functions automatically using ML and NLP methods. This said, future research may consider ML pipelines including also the formal coding of social behaviors and interactions in the audio files [58] collected by the passive sampling of participants' language use in everyday life (e.g., via the EAR). Finally, our work is based on a single study. In particular, the transcripts we considered were in German. Therefore, future research needs to investigate the generalizability of our results, including the use of different languages.

5. Conclusion

Naturalistic observation methods and unobtrusive technology allow collecting young adults' everyday conversations to study the production of reminiscence and its functions beyond existing techniques used to investigate autobiographical memories. We have provided a quantitative analysis of the occurrence of reminiscence and its functions in young adults' everyday life conversations using the corpus of transcriptions of their utterances collected with the EAR technology. Using NLP and ML, we automatically detected reminiscence and its negative functions by considering the transcripts of young adults' everyday conversations. Our approach replicates previous attempts with the detection of older adults' reminiscence in everyday life and it improves the performance of existing benchmarks. Then, we provided actionable recommendations on how to use our NLP and ML methodology to design digital health interventions that foster young adults' well-being by promoting self-reflection on their daily reminiscence activities. We also commented on the use of value sensitive design to address young adults' concerns and wishes for digital health technology. In summary, our study's quantitative approach and results open up new avenues for investigating young adults' production of autobiographical memories in their everyday life using machine learning and natural language processing methods. Further, it informs the design of digital health interventions managing young adults' well-being by promoting self-reflection on reminiscing and proving ways to cope with the emergence of reminiscence events that are detrimental to their physical and mental health.

Data availability statement

The raw/processed data required to reproduce the above findings cannot be shared at this time due to legal/ethical reasons.

CRediT authorship contribution statement

Andrea Ferrario: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Resources, Project administration, Methodology, Formal analysis, Conceptualization. **Burcu Demiray:** Writing – review & editing, Visualization, Resources, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

B.D. acknowledges the support of the Velux Stiftung (Project no. 917) and thanks Chiara Carrel, Sofie Künzle and Elia von Moos who manually coded the dataset for reminiscence and its functions.

Appendix

Table A1
POS-tags

Variable	Description
INTJ	Interjection
PUNCT	Punctuation
VERB	Verb
PART	Particle
NOUN	Noun
ADP	Adposition
DET	Determiner
ADJ	Adjective
PRON	Pronoun
ADV	Adverb
NUM	Numeral
AUX	Auxiliary
CCONJ	Coordinating conjunction
PROPN	Proper noun
INTJ	Interjection
SCONJ	Subordinating conjunction
SYM	Symbol

Table A2
Grids for Hyperparameter Tuning

Pipeline element	Hyperparameters	Grid
SMOTE	number of neighbors	[1,5,9,13]
BOW features	n-grams	[(1, 1), (1, 2)]
	stopwords	[`stopwords`,[1] None]
POS features	n-grams	[(1, 1), (1, 2)]
RFE algorithm	number of features to select (Experiment I)	[10,20,30]
	number of features to select (Experiment II)	[5,10,15,20,30]
	number of features to reduce at each step	[0.25, 0.5]
ML models (XGB, LIGHT)	number of trees	[10, 50, 100, 200, 300, 400]
	maximum tree depth	[1,3,5,10]
	learning rate	[0.001, 0.01, 0.1, 1.0]
ML models (RF)	number of trees	[10, 50, 100, 200, 300, 400]
	maximum tree depth	[1,3,5,10]
Voting classifier (4 models)	Ensemble weights ω_i	[5, 15, 25, 35, 45, 55, 65, 75, 85, 95][2]
Voting classifier (3 models)	Ensemble weights ω_i	[10, 20, 30, 33, 34, 40, 50, 60, 70, 80][2]
Voting classifier (2 models)	Ensemble weights ω_i	[10, 20, 30, 40, 50, 60, 70, 80, 90][2]

[1] Stopwords is the list of German stopwords in the Python library spacy.
[2] The *N*-th Cartesian product of the grid is computed and only the *N*-tuples summing to 100 are kept for tuning, where *N* is the number of models in the ensemble.

References

[1] S. Bluck, L.J. Levine, Reminiscence as autobiographical memory: a catalyst for reminiscence theory development, Ageing Soc 18 (1998) 185–208, https://doi.org/10.1017/S0144686X98006862.
[2] R.N. Butler, The life review: an interpretation of reminiscence in the aged, Psychiatry 26 (1963) 65–76.
[3] J.D. Webster, M.E. McCall, Reminiscence functions across adulthood: a replication and extension, J. Adult Dev. 6 (1999) 73–85.
[4] G.J. Westerhof, E. Bohlmeijer, J.D. Webster, Reminiscence and mental health: a review of recent progress in theory, research and interventions, Ageing Soc. 30 (2010) 697–721.
[5] B. Demiray, M. Mischler, M. Martin, Reminiscence in everyday conversations: a naturalistic observation study of older adults, J. Gerontol. B Psychol. Sci. Soc. Sci. 74 (2019) 745–755.
[6] J.D. Webster, O. Gould, Reminiscence and vivid personal memories across adulthood, Int. J. Aging Hum. Dev. 64 (2007) 149–170.
[7] K.C. McLean, M. Pasupathi, J.L. Pals, Selves creating stories creating selves: a process model of self-development, Pers. Soc. Psychol. Rev. 11 (2007) 262–278.
[8] D.T. Hyland, A.M. Ackerman, Reminiscence and autobiographical memory in the study of the personal past, J. Gerontol. 43 (1988) P35–P39.
[9] S.B. Merriam, L.H. Cross, Adulthood and reminiscence: a descriptive study, Educ. Gerontol. 8 (1982) 275–290.
[10] M. Pasupathi, L.L. Carstensen, Age and emotional experience during mutual reminiscing, Psychol. Aging 18 (2003) 430–442.
[11] M. Romaniuk, J.G. Romaniuk, Life events and reminiscence: a comparison of the memories of young and old adults, Imagin., Cognit. Pers. 2 (1982) 125–136.
[12] J.D. Webster, Adult age differences in reminiscence functions. The art and science of reminiscing: Theory, research, methods, and applications, Taylor & Francis, Philadelphia, PA, US, 1995, pp. 89–102.
[13] D.J. Hallford, S. Hardgrove, M. Sanam, S. Oliveira, M. Pilon, T. Duran, Remembering for resilience: brief cognitive-reminiscence therapy improves psychological resources and mental well-being in young adults, Appl Psychol Health Well Being 14 (2022) 1004–1021.
[14] D. Hallford, D. Mellor, Reminiscence-based therapies for depression: should they be used only with older adults? Clin. Psychol. Sci. Pract. 20 (2013) 452–468.

[15] T. Habermas, C. Köber, Autobiographical reasoning is constitutive for narrative identity: the role of the life story for personal continuity. The oxford handbook of identity development, Oxford University Press, New York, NY, US, 2015, pp. 149–165.
[16] D.P. McAdams, B.D. Olson, Personality development: continuity and change over the life course, Annu. Rev. Psychol. 61 (2010) 517–542.
[17] K.L. Graham, R.L. Rahm-Knigge, B.T. Conner, Profiling reminiscers: using a self-report measure of frequency and functions of reminiscence to identify individual patterns of retrieval, J. Adult Dev. 27 (2020) 170–180.
[18] J.D. Webster, Construction and validation of the reminiscence functions scale, J. Gerontol. 48 (1993) P256–P262.
[19] J.D. Webster, The reminiscence functions scale: a replication, Int. J. Aging Hum. Dev. 44 (1997) 137–148.
[20] P. Cappeliez, R. Lavallee, Functions of reminiscence in later life as viewed by young and old adults, Can. J. Aging 20 (2001).
[21] N. O'Rourke, P. Cappeliez, A. Claxton, Functions of Reminiscence and the Psychological Well-Being of Young-Old and Older Adults over Time, Aging & mental health, 2011.
[22] N. O'Rourke, D.B. King, P. Cappeliez, Reminiscence functions over time: consistency of self functions and variation of prosocial functions, Memory 25 (2017) 403–411.
[23] P. Cappeliez, N. O'Rourke, Empirical validation of a model of reminiscence and health in later life, J. Gerontol. B Psychol. Sci. Soc. Sci. 61 (2006) P237–P244.
[24] M.R. Mehl, J.W. Pennebaker, D.M. Crow, J. Dabbs, J.H. Price, The Electronically Activated Recorder (EAR): a device for sampling naturalistic daily activities and conversations, Behav. Res. Methods Instrum. Comput. 33 (2001) 517–523.
[25] M.R. Mehl, S.E. Holleran, An empirical analysis of the obtrusiveness of and participants' compliance with the electronically activated recorder (EAR), Eur. J. Psychol. Assess. 23 (2007) 248–257.
[26] B. Demiray, M. Luo, A. Tejeda-Padron, M.R. Mehl, Sounds of healthy aging: assessing everyday social and cognitive activity from ecologically sampled ambient audio data, in: P.L. Hill, M. Allemand (Eds.), Personality and Healthy Aging in Adulthood: New Directions and Techniques [Internet], Springer International Publishing, Cham, 2020, https://doi.org/10.1007/978-3-030-32053-9_8 [cited 2020 Mar 23]. pp. 111–32.
[27] M.R. Mehl, The Electronically Activated Recorder (EAR): a method for the naturalistic observation of daily social behavior, Curr. Dir. Psychol. Sci. 26 (2017) 184–190.
[28] A. Ferrario, B. Demiray, K. Yordanova, M. Luo, M. Martin, Social reminiscence in older adults' everyday conversations: automated detection using natural language processing and machine learning, J. Med. Internet Res. 22 (2020), e19133.
[29] A. Ferrario, M. Luo, A.J. Polsinelli, S.A. Moseley, M.R. Mehl, K. Yordanova, et al., Predicting working memory in healthy older adults using real-life language and social context information: a machine learning approach, JMIR Aging 5 (2022), e28333.
[30] T. Stoev, A. Ferrario, B. Demiray, M. Luo, M. Martin, K. Yordanova, Coping with imbalanced data in the automated detection of reminiscence from everyday life conversations of older adults, IEEE Access 9 (2021) 116540–116551.
[31] K.Y. Yordanova, B. Demiray, M.R. Mehl, M. Martin, Automatic detection of everyday social behaviours and environments from verbatim transcripts of daily conversations. proceedings of IEEE international conference on pervasive computing and communications [Internet], IEEE, Kyoto, Japan, 2019 [cited 2020 Mar 5]. pp. 242–51. Available from: https://ieeexplore.ieee.org/document/8767403.
[32] M. Luo, R. Debelak, G. Schneider, M. Martin, B. Demiray, With a little help from familiar interlocutors: real-world language use in young and older adults, Aging Ment. Health 25 (2021) 2310–2319.
[33] A. Macbeth, M.R. Bruni, B. De La Cruz, J.A. Erens, N. Atagi, M.L. Robbins, et al., Using the electronically activated recorder (EAR) to capture the day-to-day linguistic experiences of young adults, Collabra: Psychology 8 (2022), 36310.
[34] B. Demiray, M.R. Mehl, M. Martin, Conversational time travel: evidence of a retrospective bias in real life conversations, Front. Psychol. 9 (2018). Available from: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02160.
[35] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et al. (Eds.), Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., 2017 [cited 2020 Jan 3]. pp. 4765–4774. Available from: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.
[36] M.L. Robbins, Practical suggestions for legal and ethical concerns with social environment sampling methods, Soc. Psychol. Personal. Sci. 8 (2017) 573–580.
[37] M.A. Conway, J.A. Singer, A. Tagini, The self and autobiographical memory: correspondence and coherence, Soc. Cognit. 22 (2004) 491–529.
[38] D. Jurafsky, J.H. Martin, Speech and Language Processing, second ed., Prentice Hall, 2014.
[39] C. Manning, H. Schütze, Foundations of Statistical Natural Language Processing, MIT Press, 1999. Available from: https://nlp.stanford.edu/fsnlp/.
[40] Z.S. Harris, Distributional structure, Word 10 (1954) 146–162.
[41] A. Ferrario, M. Naegelin, The Art of natural language processing: Classical, modern and contemporary approaches to text document classification, Social Science Research Network, Rochester, NY, 2020. Mar. Report No.: ID 3547887. Available from: https://papers.ssrn.com/abstract=3547887.
[42] V. López, A. Fernández, J.G. Moreno-Torres, F. Herrera, Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics, Expert Syst. Appl. 39 (2012) 6585–6608.
[43] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic Minority Over-sampling TEchnique, JAIR 16 (2002) 321–357.
[44] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (2002) 389–422.
[45] T. Hastie, R. Tibshirani, J. Friedman, Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2009. Available from: https://web.stanford.edu/~hastie/ElemStatLearn/.
[46] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16 (2016) 785–794.
[47] A. Ferrario, R. Hämmerli, On boosting: Theory and applications, social science research network, Rochester, NY, 2019 Jun. Report No.: ID 3402687. Available from: https://papers.ssrn.com/abstract=3402687.
[48] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, et al., LightGBM: a highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017. Available from: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.
[49] D. Ruta, B. Gabrys, Classifier selection for majority voting, Inf. Fusion 6 (2005) 63–81.
[50] T. Chen, J. Xu, H. Ying, X. Chen, R. Feng, X. Fang, et al., Prediction of extubation failure for intensive care unit patients using light gradient boosting machine, IEEE Access 7 (2019) 150960–150968.
[51] J. Zhang, T. Jun, J. Frank, S. Nirenberg, P. Kovatch, K. Huang, Prediction of individual COVID-19 diagnosis using baseline demographics and lab data, Sci. Rep. 11 (2021), 13913.
[52] M.J. Ariza-Garzón, J. Arroyo, A. Caparrini, M.-J. Segovia-Vargas, Explainability of a machine learning granting scoring model in peer-to-peer lending, IEEE Access 8 (2020) 64873–64890.
[53] M. Naegelin, R.P. Weibel, J.I. Kerr, V.R. Schinazi, R. La Marca, F. von Wangenheim, et al., An interpretable machine learning approach to multimodal stress detection in a simulated office environment, J. Biomed. Inf. 139 (2023), 104299.
[54] M.T. Ribeiro, S. Singh, C. Guestrin, "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA, 2016, https://doi.org/10.1145/2939672.2939778, pp. 1135–1144.
[55] L.S. Shapley, Notes on the N-Person Game — II: the Value of an N-Person Game, RAND Corporation, 1951 Aug. Available from: https://www.rand.org/pubs/research_memoranda/RM0670.html.
[56] B. Efron, Bootstrap methods: another look at the jackknife, Ann. Stat. 7 (1979) 1–26, https://doi.org/10.1007/978-1-4612-4380-9_41.
[57] B. Colombo, S. Balzarotti, A. Greenwood, Using a reminiscence-based approach to investigate the cognitive reserve of a healthy aging population, Clin. Gerontol. 42 (2019) 408–420.
[58] T. Stivers, Coding social interaction: a heretical approach in conversation analysis? Res. Lang. Soc. Interact. 48 (2015) 1–19.

[59] A.F. Markus, J.A. Kors, P.R. Rijnbeek, The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies, J. Biomed. Inf. 113 (2021), 103655.
[60] M. Pinquart, S. Forstmeier, Effects of Reminiscence Interventions on Psychosocial Outcomes: a Meta-Analysis, Centre for Reviews and Dissemination, UK, 2012. Available from: https://www.ncbi.nlm.nih.gov/books/NBK100209/.
[61] J.D. Webster, E.T. Bohlmeijer, G.J. Westerhof, Mapping the future of reminiscence: A conceptual guide for research and practice: Research on aging (2010). Available from: https://journals.sagepub.com/doi/10.1177/0164027510364122.
[62] Hallford D.J. Pilon M.F, S. Hardgrove, M. Sanam, S. Oliveira, G.C. Karantzas, D. Tyler, et al., Remembering for relationships: brief cognitive-reminiscence therapy improves young adults' perceptions about self and others in social relationships, J. Soc. Pers. Relat. 40 (2023) 955–975.
[63] D.J. Hallford, D. Mellor, Brief reminiscence activities improve state well-being and self-concept in young adults: a randomised controlled experiment, Memory 24 (2016) 1311–1320.
[64] A. Bejan, R. Gündogdu, K. Butz, N. Müller, C. Kunze, P. König, Using multimedia information and communication technology (ICT) to provide added value to reminiscence therapy for people with dementia, Z. Gerontol. Geriatr. 51 (2018) 9–15.
[65] A. Lazar, H. Thompson, G. Demiris, A systematic review of the use of technology for reminiscence therapy, Health Educ. Behav. 41 (2014) 51S–61S.
[66] S.T. Peesapati, V. Schwanda, J. Schultz, M. Lepage, S. Jeong, D. Cosley, Pensieve: supporting everyday reminiscence. Proceedings of the SIGCHI conference on human factors in computing systems, ACM, Atlanta Georgia USA, 2010, pp. 2027–36. Available from: https://dl.acm.org/doi/10.1145/1753326.1753635.
[67] S. Garrido, C. Millington, D. Cheers, K. Boydell, E. Schubert, T. Meade, et al., What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people, Front. Psychiatr. (2019), 10. Available from: https://www.frontiersin.org/articles/10.3389/fpsyt.2019.00759.
[68] A. Baumel, T. Fleming, S.M. Schueller, Digital micro interventions for behavioral and mental health gains: core components and conceptualization of digital micro intervention care, J. Med. Internet Res. 22 (2020), e20631.
[69] F. Wahle, T. Kowatsch, E. Fleisch, M. Rufer, S. Weidt, Mobile sensing and support for people with depression: a pilot trial in the wild, JMIR mHealth and uHealth 4 (2016), e5960.
[70] J.I. Kerr, M. Naegelin, M. Benk, F v Wangenheim, E. Meins, E. Viganò, et al., Investigating employees' concerns and wishes regarding digital stress management interventions with value sensitive design: mixed methods study, J. Med. Internet Res. 25 (2023), e44131.
[71] A. Ferrario, M. Loi, How Explainability Contributes to Trust in AI. 2022 ACM Conference on Fairness, Accountability, and Transparency, Association for Computing Machinery, New York, NY, USA, 2022, https://doi.org/10.1145/3531146.3533202, pp. 1457–1466.
[72] N. Dragano, T. Lunau, Technostress at work and mental health: concepts and research results, Curr. Opin. Psychiatr. 33 (2020) 407–413.
[73] B. Friedman, P.H. Kahn, A. Borning, A. Huldtgren, Value sensitive design and information systems, in: N. Doorn, D. Schuurbiers, I. van de Poel, M.E. Gorman (Eds.), Early Engagement and New Technologies: Opening up the Laboratory, Springer Netherlands, Dordrecht, 2013, https://doi.org/10.1007/978-94-007-7844-3_4, pp. 55–95.
[74] I. Maathuis, M. Niezen, D. Buitenweg, I.L. Bongers, C. van Nieuwenhuizen, Exploring human values in the design of a web-based QoL-instrument for people with mental health problems: a value sensitive design approach, Sci. Eng. Ethics 26 (2020) 871–898.
[75] M. Mueller, O. Heger, B. Niehaves, Investigating ethical design requirements for digitalized Healthcare support [Internet], The Case of Ambulatory Physiotherapeutic Assistance Systems (2018). Available from: http://hdl.handle.net/10125/50260.
[76] A.J. Thirunavukarasu, D.S.J. Ting, K. Elangovan, L. Gutierrez, T.F. Tan, D.S.W. Ting, Large language models in medicine, Nat Med 29 (2023) 1930–1940.