

AI-enabled Language Processing of Adolescents' and Young Adults' Challenges and Aspirations (ALPACA)



Dr. Christina Haag & Dr. David Bürgin



Universität
Zürich ^{UZH}

UZH
Institute for
Implementation
Science in
Health Care



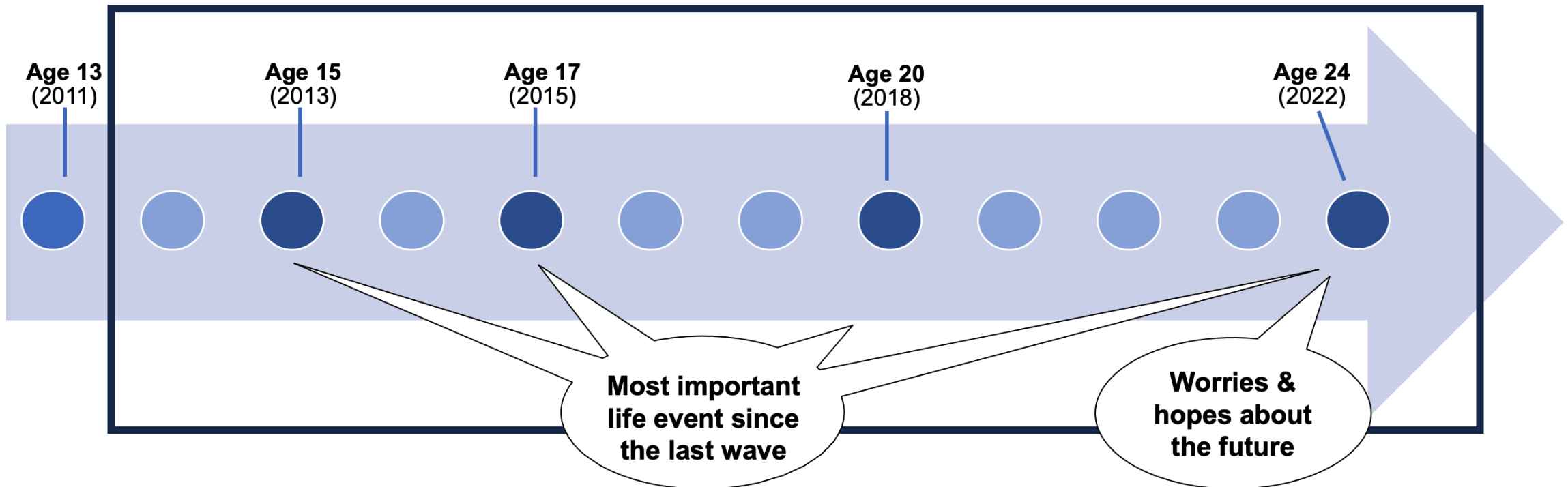
**JACOBS
CENTER**

Z-proso
Zürcher Projekt zur sozialen Entwicklung
von der Kindheit ins Erwachsenenalter

Seed grant from the

UZH PRC
Population Research Center

Zurich Project on the Social Development from Childhood to Adulthood (z-proso)



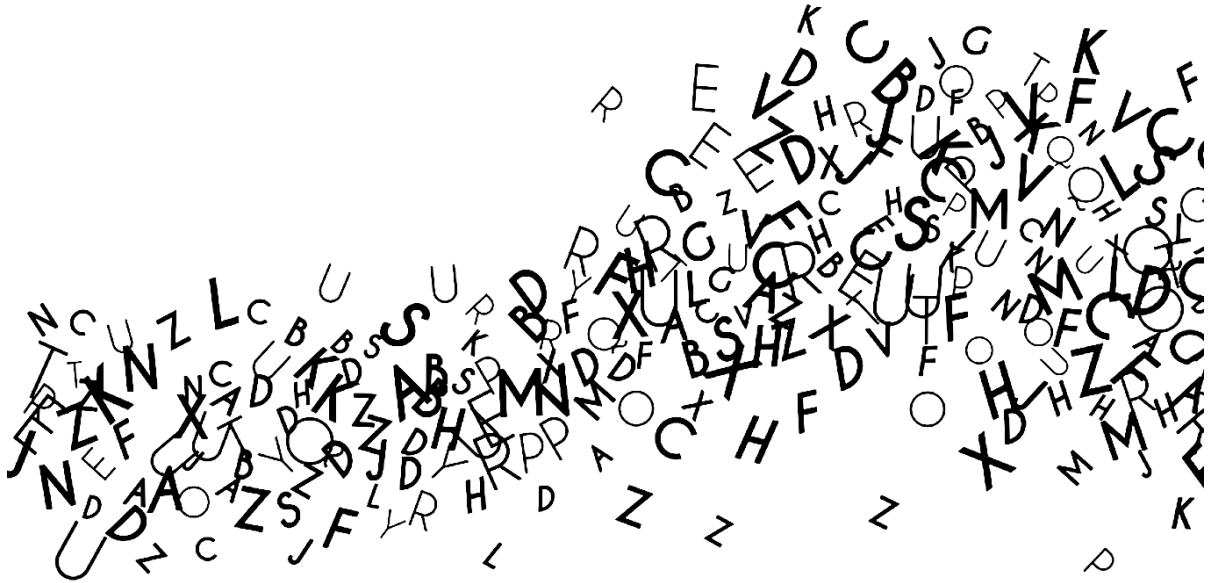


Study Aims

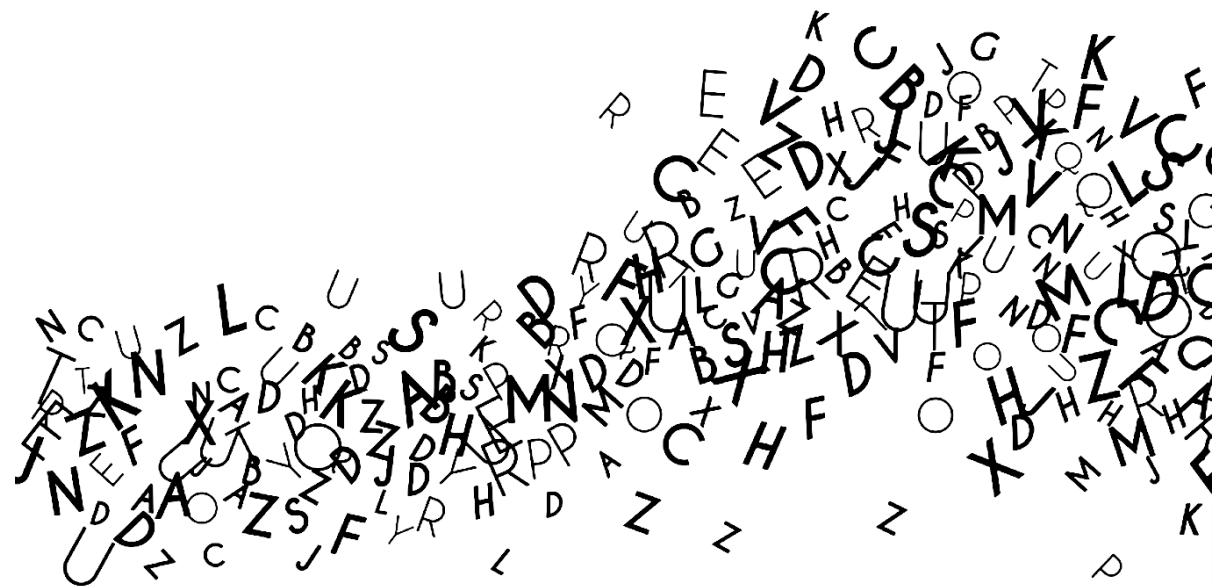
Aim 1: Identifying major life events from adolescence to young adulthood by

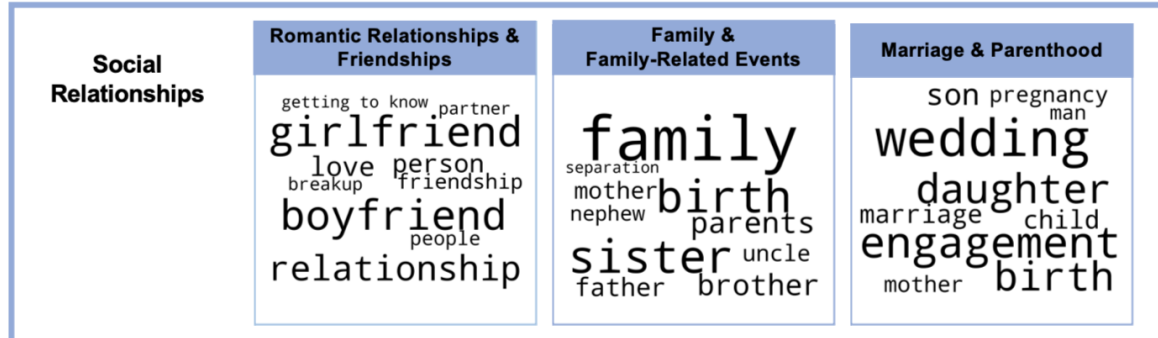
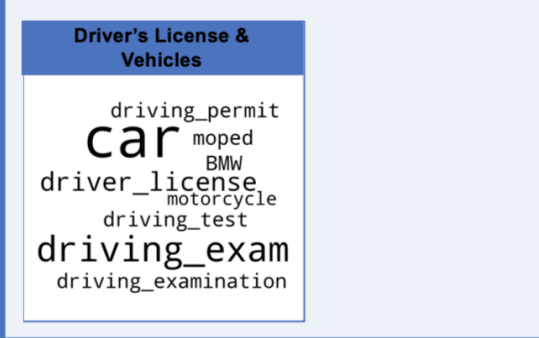
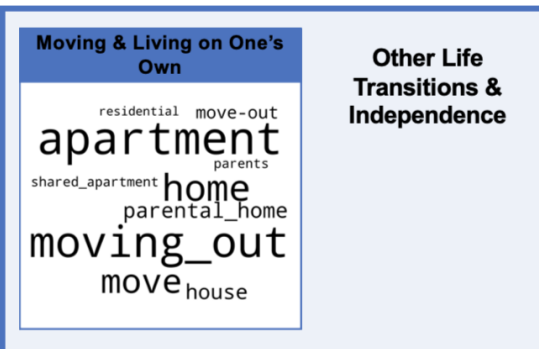
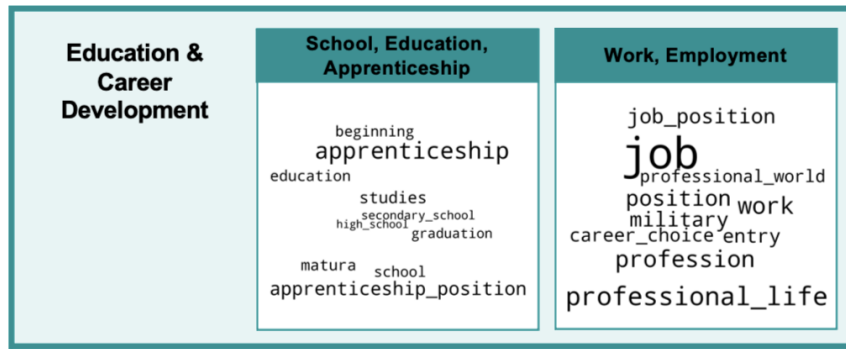
- identifying key topics
- emotional valence
- transitions over time

Aim 2: Identifying key themes in young adults' worries and hopes for the future and training a text-based mental wellbeing classifier

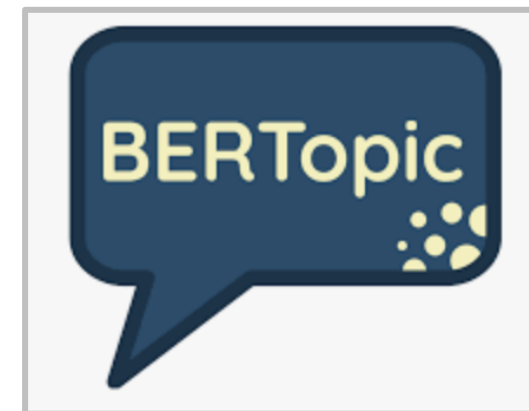


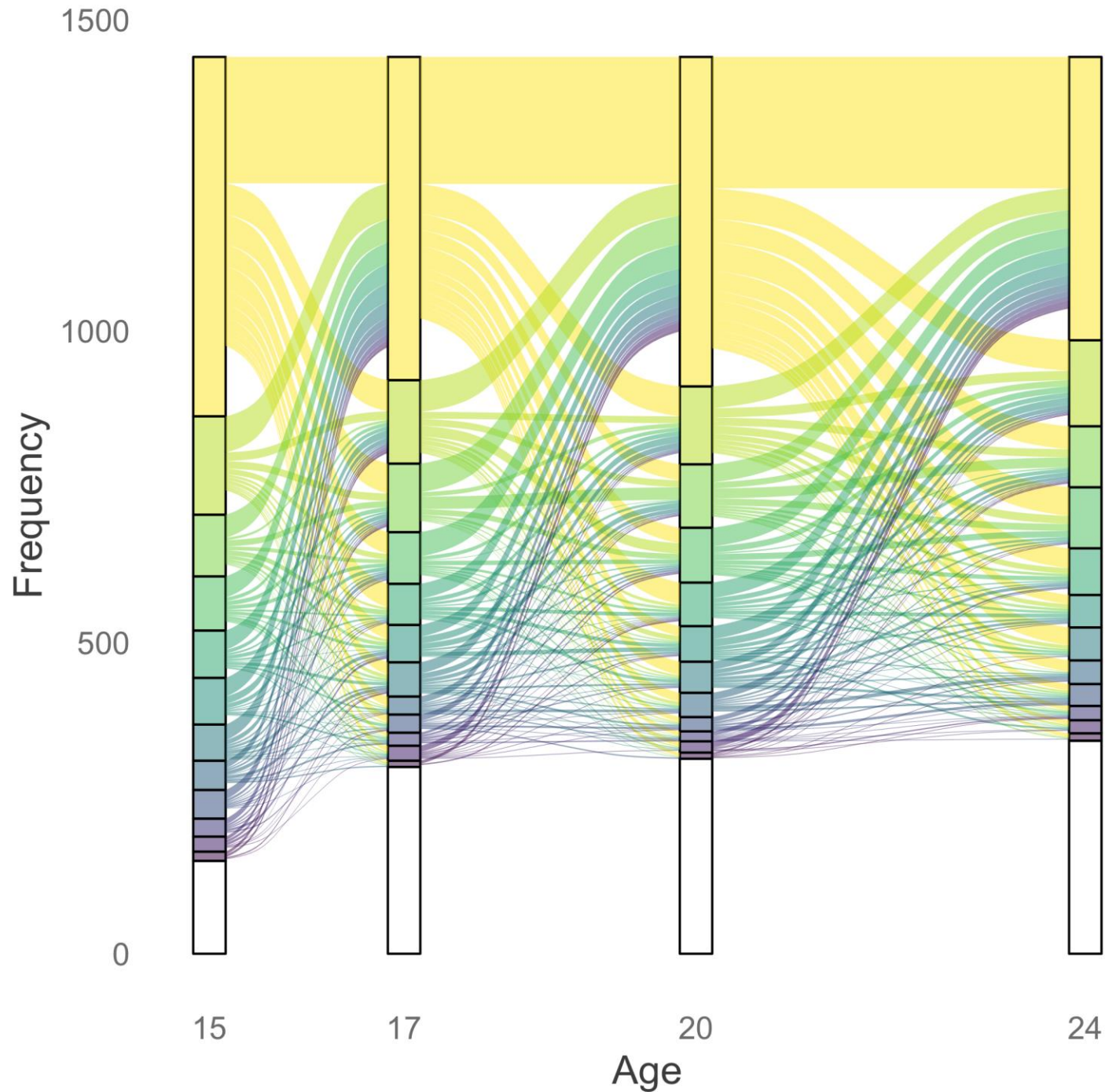
1. Converting text data to numerical format via an embedding model	Language model: <i>'paraphrase-multilingual-MiniLM-L12-v2'</i>
2. Reducing data dimensionality	UMAP algorithm
3. 'Hard' clustering: Each text entry is categorised into precisely one topic or marked as an outlier	HDBSCAN algorithm
4. Splitting text entries into words ('tokenization')	CountTokenizer (scikit-learn)
5. Lemmatization, part-of-speech tagging	spaCy pipeline <i>'de_dep_news_trf'</i>
6. Model fine-tuning to avoid topics consisting of highly redundant keywords	Maximal Marginal Relevance criterion
7. Computing importance scores for each word in a cluster	c-TF-IDF algorithm
8. Merging topics with overlapping content based on previous manual review	BERTopic's <i>'merge_topics'</i> command
9. 'Soft' clustering assigns probabilities to the entries, allowing membership in overlapping clusters	HDBSCAN algorithm
10. Updating Topic Representations: Texts are classified under specific topics or identified as outliers	BERTopic's <i>'update_topics'</i> command
11. Manual check and correction if necessary	Manual Review





5,708 text-entries,
1,442 participants
4 timepoints






Topic

- School, Education, & Apprenticeship
- Romantic Relationships & Friendship
- Mental Health, Development, & Change
- Travelling, Vaccation, & Staying Abroad
- Work & Employment
- Family & family-related Events
- Moving & Living on One's Own
- Stressful Life Events & Loss
- Sports
- Celebrations & Going out
- Marriage & Parenthood
- Driver's Licence & Vehicles
- NA

Classifier training to identify emotional valence



 Hugging Face

 `google-bert/bert-base-german-dbmdz-uncased`

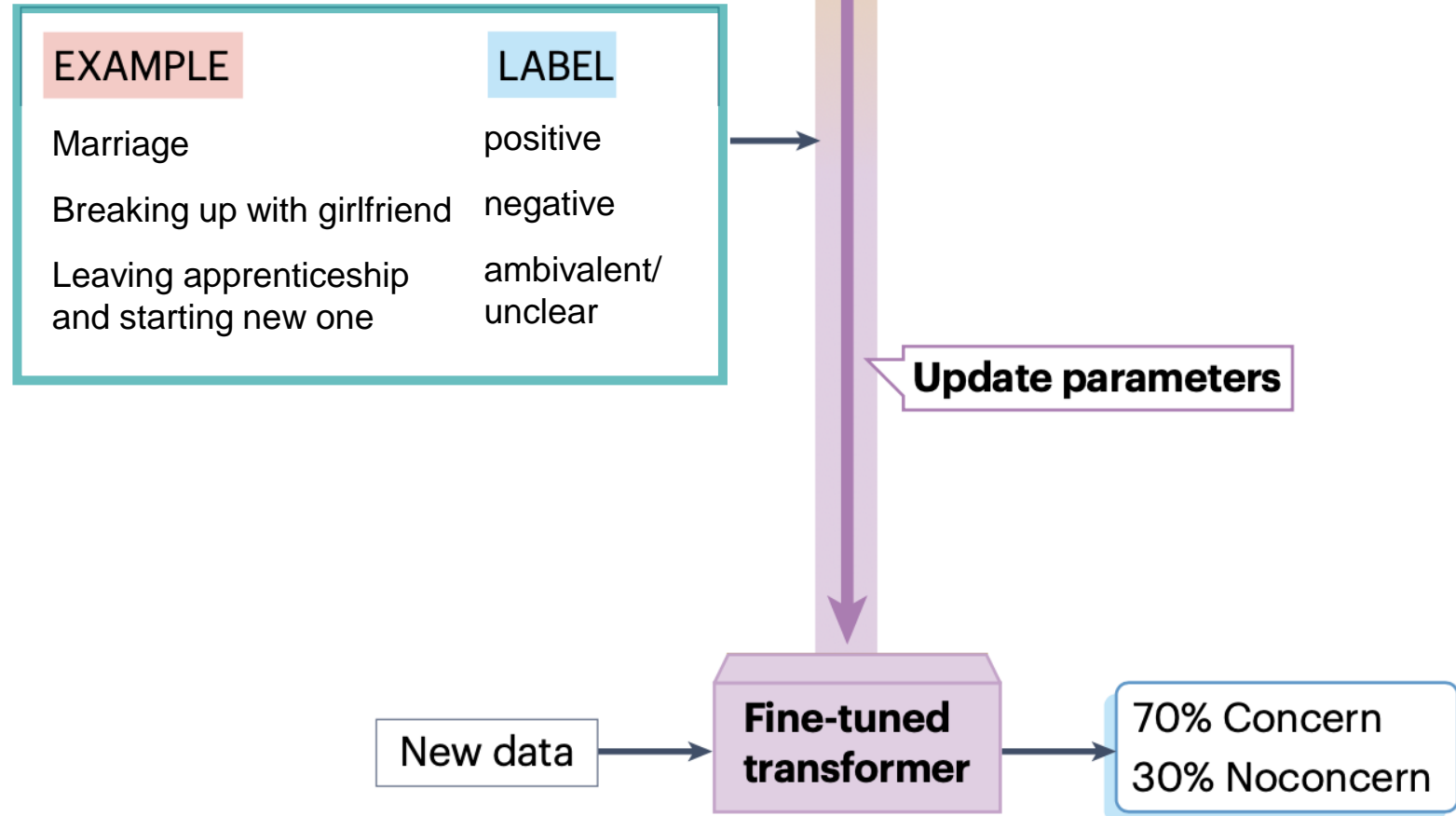
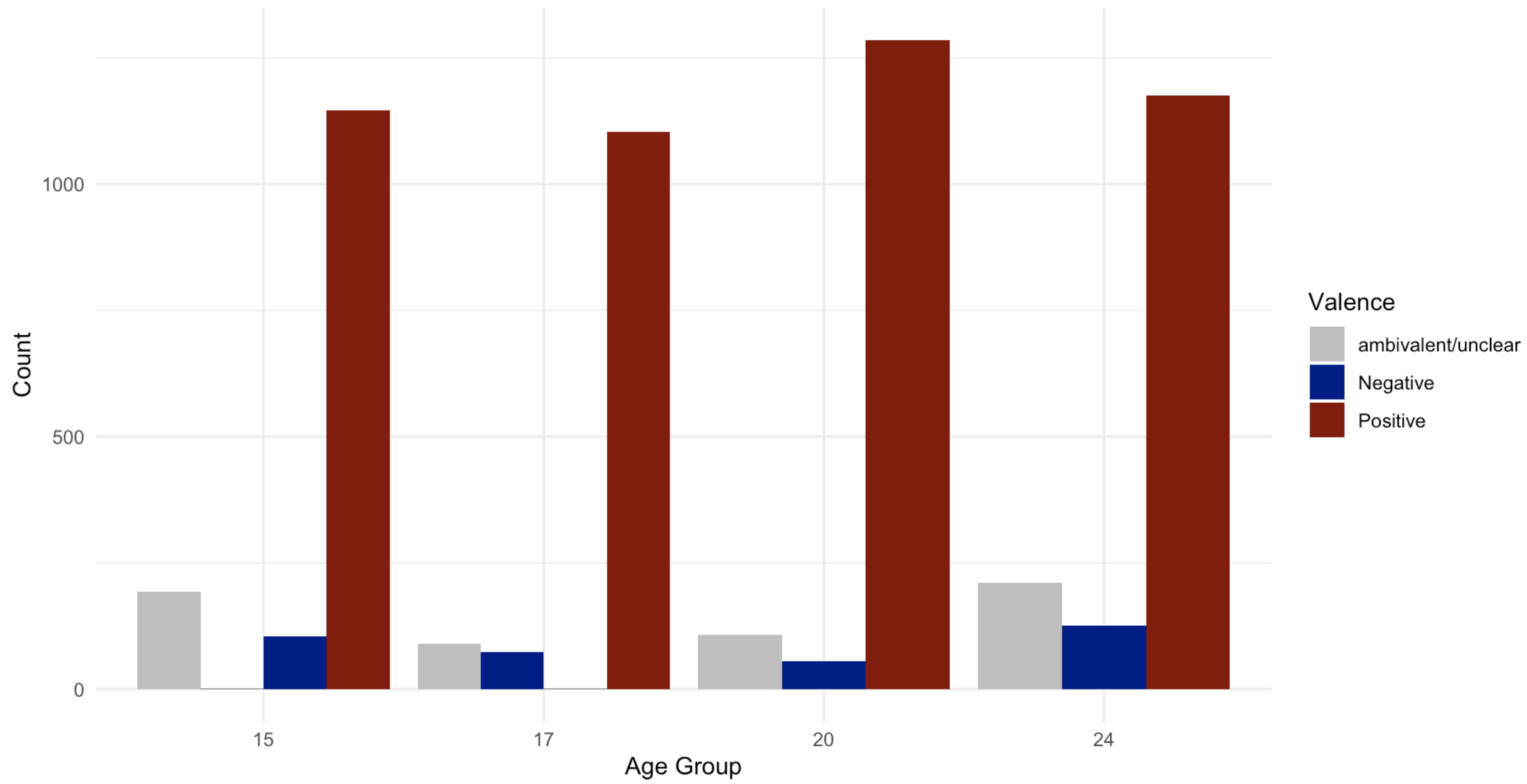
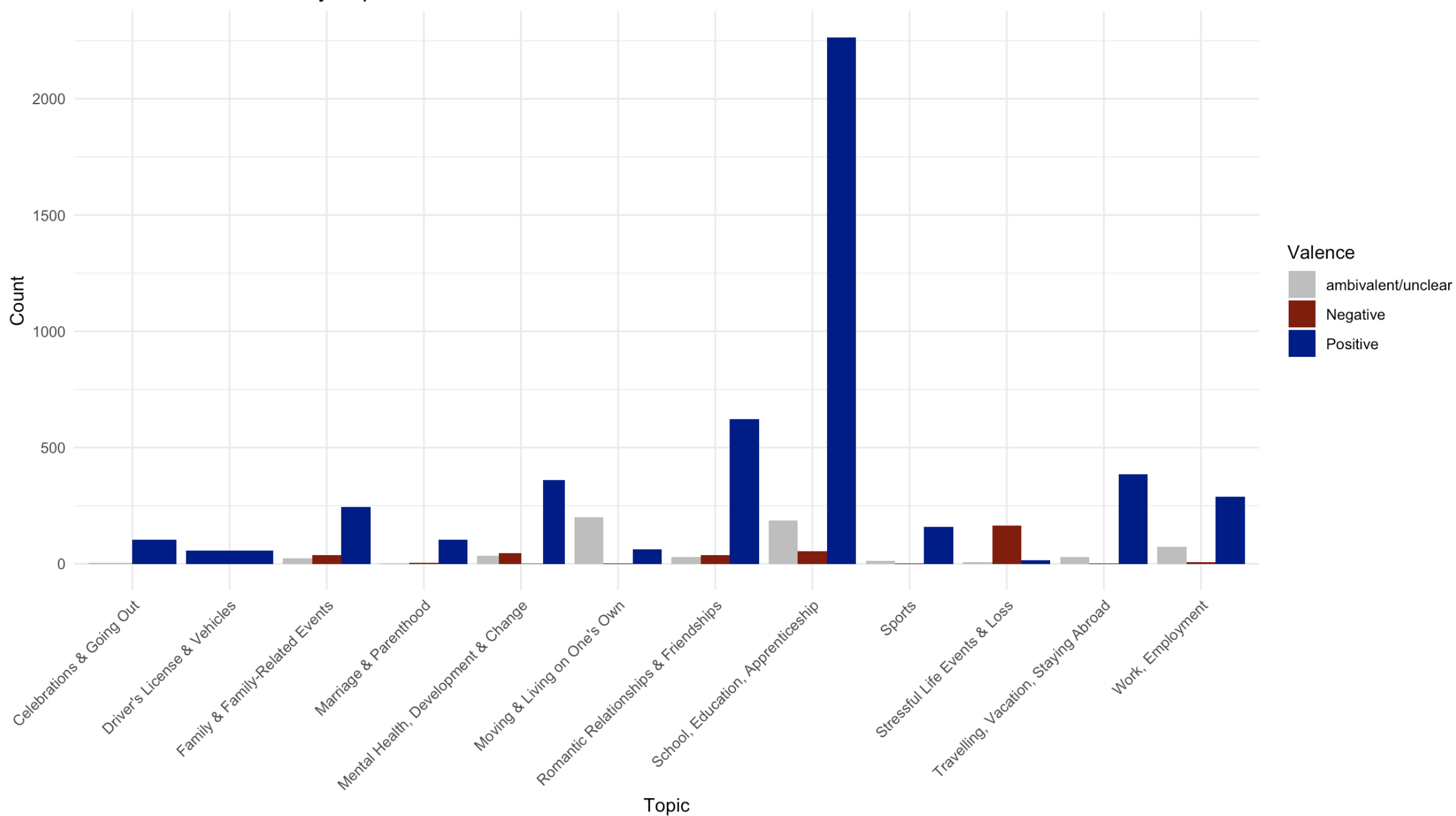


Image sourced from Demszky et al. (2023), published in Nature Reviews Psychology

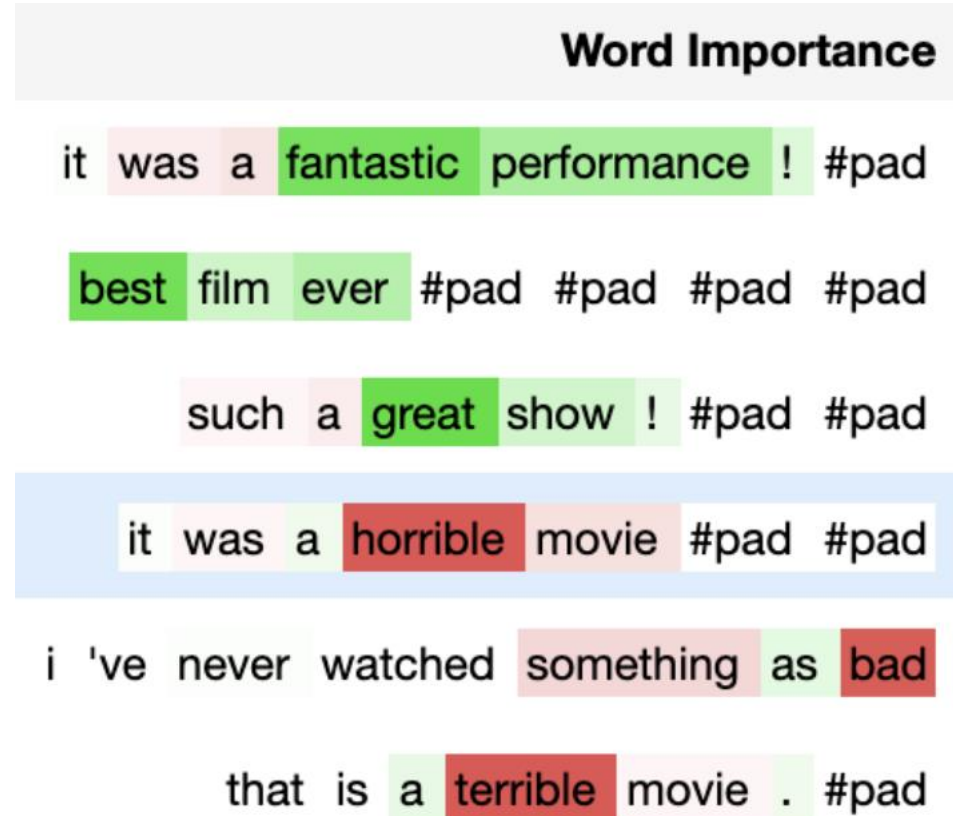
Distribution of Valence by Age Group



Distribution of Valence by Topic



To be completed: Explainable AI Analysis



Integrating traditional quantitative methods with automated text analysis can lead to more accurate and personalised insights.

Dr. David Bürgin,
Dr. Christina Haag

&

Lynn Alison Büeler, MSc



Prof. Lilly Shanahan &
Prof. Viktor von Wyl

Dr. Denis Ribeaud &
Prof. Manuel Eisner

Dr. Laura Bechtiger &
Dr. Clarissa Janousch



Universität
Zürich ^{UZH}

UZH PRC
Population Research Center



UZH
Institute for
Implementation
Science in
Health Care

