Personality as Search:
Automated Big Five Assessment
with BERT (ColB5ERT)

( )

Ricardo Primi

University of São Francisco & Edulab21, Ayrton Senna Institute, Brazil





R2D2 Powered by GPT8



## Research question

Since digital footprints provide immense sample of people's behavior (Matz, 2025)

Can we measure a person's big five traits (but also values, moral foundations, interests) from their written texts?

Spontaneous texts: Posts, written messages on social media.

Elicited texts: Autobiographies, written responses to structured questions about personality traits.

Can be a continuous monitor for assessment

A new general method for assessment.

## **Method Overview**

Framing self report assessment as Computer Science Information Search

Training a BERT model using Sentence Transformers / Colbert = ColB5ert

Using the model to score texts

Studying the statistical distributions automated measures of B5 and validity

# Personality assessment as search

## Basic Elements of Information Search in Computer Science

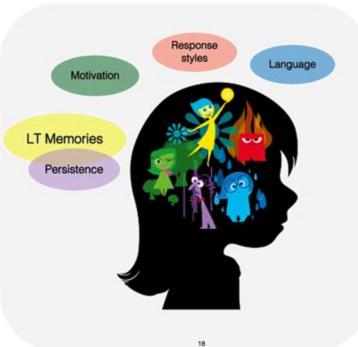
Query: The user's input or question expressed in a natural or structured language.

Document Corpus: The collection of documents or data sources to be searched.

Search algorithms by semantic similarity: The methods used to match queries with relevant documents based on words (past) or more recently contextual word vectors (embeddings)

Relevance Scoring: Ranking documents based on their semantic similarity to the query.

Retrieval: Fetching the most relevant documents based on the relevance scores



When confronted with a problem:

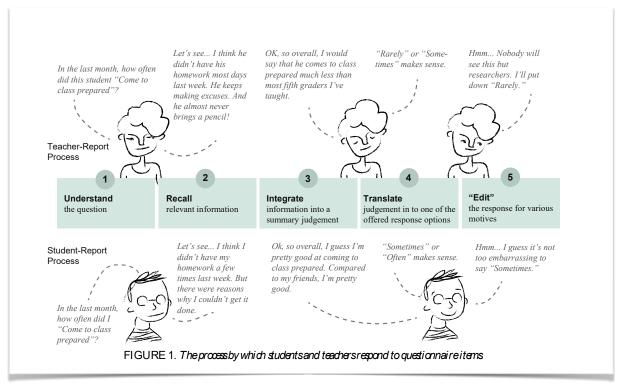
I give up easily?

I do more?

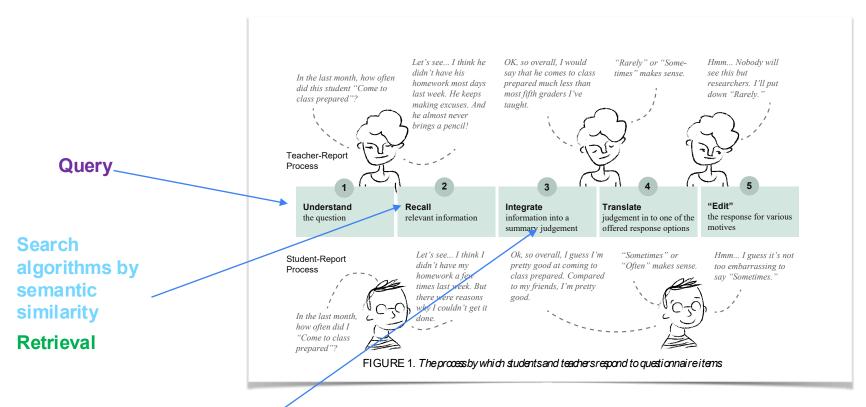
I remain interested?

## The cognitive process of Self-report

Duckworth AL, Yeager DS. Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes. Educ Res. 2015 May;44(4):237-251. doi: 10.3102/0013189X15584327. PMID: 27134288; PMCID: PMC4849415.

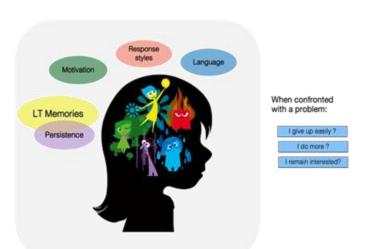


## Elements of Search on Self-report



**Relevance Scoring** 

## Personality assessment and Information Search



**Statements/items** of a test are like questions/queries.

The life story stored in long-term memory consists of passages from a comprehensive document that represents the person's complete autobiography (person life's story corpus).

The Likert response to an item is a **relevance score** of the query (items) in signaling that relevant passage was retrieved!

A measure of personality is an aggregation of relevance scores by query category (extroversion, openness, etc.).

We rely on the person's **internal search mechanism** to find degree of similarity between items and memories



	Elements of IS	Information Search	Personality Assessment	Digital footprints (posts in FB)				
	Purpose	Find relevant documents	Score queries (itens)  items organized by a scheme (psychological theory)	Score queries (itens)				
	Query	Subject generated	Professional/research generated (limited)	Professional/research generated (potentially unlimited)				
	Document	An entire knowledge base from which we want to recover a few relevant docs	Personal History from which we want to detect if a query match relevant LT memories	All history of posts from which we want to score specific queries				
	Search process	Exhaustive	Selective filtering	Exhaustive				
	Retrieval	Retrieval of a subset with high relevance scores	Specific fact(s) / memory(ies) or experience of familiarity	Exhaustive				
-	Relevance score	Number of words or sum of words semantic similarity between queries and documents	Likert response, score on scales (how much item words were part of personal history)	Semantic similarity between queries and posts				

## Al: Transformers 101 BERT

## How to measure semantic similarity between texts (spontaneous texts and personality items)

101 on Transformers

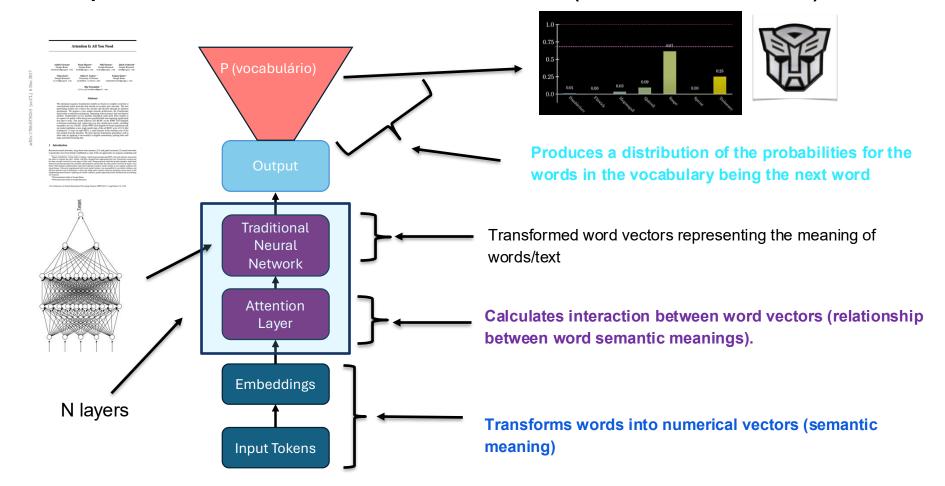
GPT: General Pré Trained Transformers

BERT: Bi-Directional Encoders from Transformers

Rich Latent Representations of texts: Contextualized

**Embeddings** 

## Simplified scheme of a Transformer (GPT and BERT)



## **Embeddings:** From Words to Meaning: How LLMs Represent Language

Words (tokens) are converted into **vectors of continuous numbers** → *embeddings* 

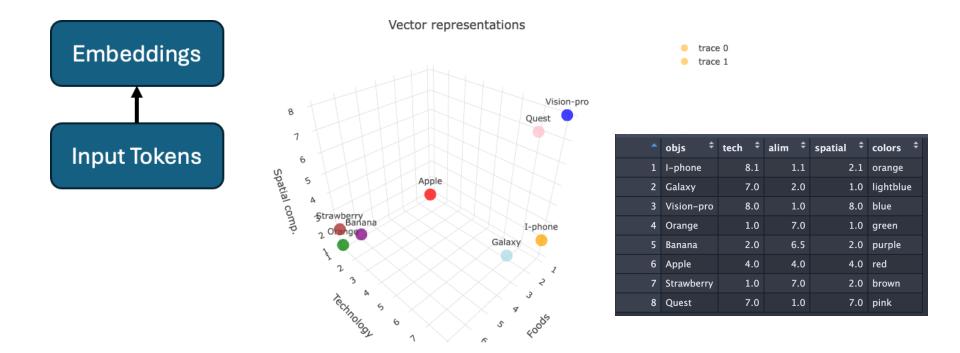
Embeddings capture **relationships** between words

Encode **semantic meaning**, not just surface form

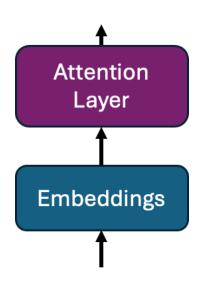
Encode **deep abstractions structures**: grammar, code, symbolic math, logic systems

Emerged from training on **billions of words** from diverse sources and Large context windows that let models analyze **meaning across long passages** 

# I bought a box of apples. I bought an apple i-phone.



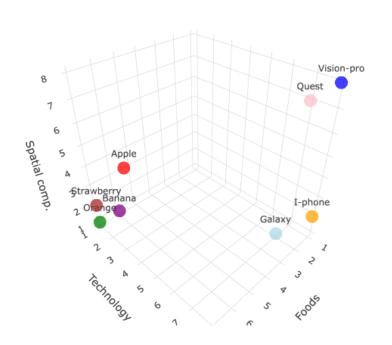
- Each word has a numerical vector of dimension *D* (GPT-4: 1536 dimensions).
- Numbers are learned such that each dimension encodes a semantic attribute.
- Similar words have similar numerical vectors, which means they are close to each other in the Ddimensional space.
- Therefore, the correlation between vectors (cosine similarity) is higher for similar words.

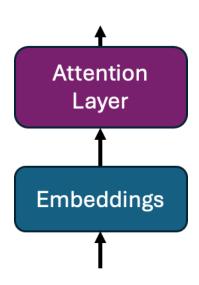


## I bought a box of apples

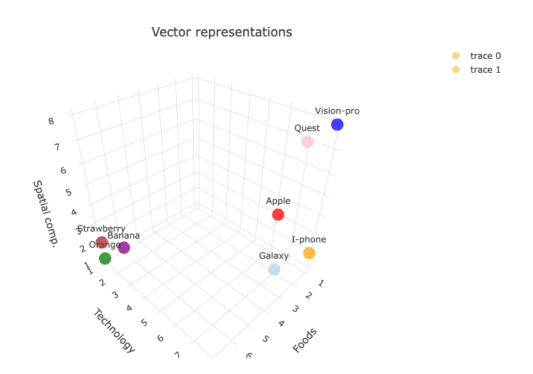


trace 0 trace 1

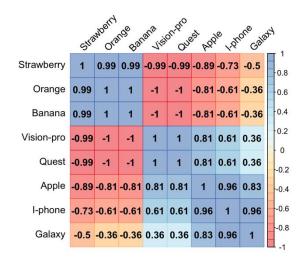




## I bought an apple i-phone.



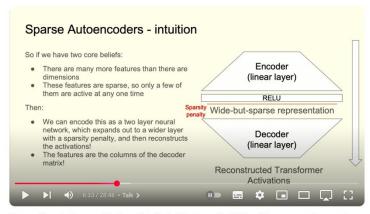
	objs ‡	tech ‡	alim ‡	spatial ‡	colors ‡
1	I-phone	8.1	1.1	2.1	orange
2	Galaxy	7.0	2.0	1.0	lightblue
3	Vision-pro	8.0	1.0	8.0	blue
4	Orange	1.0	7.0	1.0	green
5	Banana	2.0	6.5	2.0	purple
6	Apple	4.0	4.0	4.0	red
7	Strawberry	1.0	7.0	2.0	brown
8	Quest	7.0	1.0	7.0	pink



Semantic similarity between words:
Cosine similarity among embeddings (word vectors)
Correlations / Profile similarity

Sentence Semantic similarity:
Average embeddings of sentence words
Calculate similarity between resulting vectors of sentences

But ..caution on interpreting embedding dimensions as latent variables .. there are much more features than dimensions .. superposition



Hoagy Cunningham — Finding distributed features in LLMs with sparse autoencoders [TAIS 2024]

Anthropic

https://www.youtube.com/watch?v=HPLII9ZOpUQ

https://transformer-circuits.pub/2025/attribution-graphs/biology.html

## **Empirical Study**

Measuring Big Five on FB posts

## How to measure Big Five with AI?

BIG 5: Personality traits organized into five broad domains: **O**penness, **C**onscientiousness, **E**xtraversion, **A**mability and **N**egative Emotional Regulation

If we have sentence embeddings for big five personality items

Text (FB posts) embeddings

We can score semantic similarity of items-text pairs by calculating the cosine similarity

#### Goals

Fine Tune BERT to maximize detection of big five knowledge

BERT general language

Fine tunning BERT to the task of scoring Big Five

ColB5BERT / Sentence BERT

Initialize the model in its original weights

Train with a new data with a new specialized objective

#### **Building a Corpus for Personality-dependent** Natural Language Understanding and Generation

R.M.S. Ramos, G.B.S. Neto, B.B.C. Silva, D.S. Monteiro, I. Paraboni, R.F.S. Dias University of São Paulo, School of Arts, Sciences and Humanities São Paulo, Brazil

(ricellimaliva georges, stavracas harbarah claudino dasisamon irander paraboni rafarbandronidias) il gmail com

The computational treatment of human personality - both for the recognition of personality traits from text and for the generation of text so as to reflect a particular set of tests - is central to the development of NLP applications. As a means to provide a basic resource for studies of this kind, this article describes the MI corpus, a collection of controlled and free (non-topic specific) texts produced in different (e.g., referential or descriptive) communicative tasks, and accompanied by inventories of personality of their authors and additional demographics. The present discussion is mainly focused on the various components and on the data collection task itself, but proliminary results of personality recognition from test are presented in order to illustrate how the corpus data may be request. The M corpus aims to provide support for a wide range of NLP studies based on personality information and it is, to the best of our knowledge, the largest resource of this kind to be made available for research purposes in the Bracilian Portuguese language.

Keywords: Corpora, Personality, Big Five

#### 1. Introduction

In recent years, the development of so-called intelligent systems has devoted a great deal of attention to the computational treatment of human personality. This interest may be explained, among other reasons, by the practical observation that users of computer systems not only attribute human traits to the systems they interact with, but they also prefer those systems that present traits similar to their own Maircose et al., 2007).

Fundamental personality traits are consistently reflected in the language choices made by individuals when communicating. For instance, an individual with narcissistic traits might make frequent use of first-person expressions (T. 'for me', etc.). The relation between personality and natural language is the focus of a large body of work in the Psychology field, and it is perhaps best summarised by the Big Five personality factors (Coldberg, 1990) - Openness to-experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism - which are widely accepted as an adequate basis for the representation of human personality. Given its linguistic motivation, the Big Five model provides a theoretical busis for the computational treatment of personality on at least two from: the automatic recognition of personality traits from test (which is a language understanding task), and the generation of text in order to reproduce-certain personality traits of interest (which is a natural language generation (NLC) task). Knowing the personality traits of an individual (e.g., from his/her social network status updates) has many obvious applications, including staff recruitment, credit analysis etc. In addition to that, personality information may also guide the automatic generation of personalised content, the modelling of psychologically plausible virtual agents (e.g., intelligent tutors, game characters, etc.) and human-computer dialogue applications in which a high degree of realism and engagement is required. Personality-oriented language understanding and generation are considerable research challenges and, despite their

complementary nature (for example, in applications of human-computer dialogue), will usually have a common starting point: a basic resource from which we may establish mappings from linguistic features to personality traits. Based on this observation, this article presents the M one pun of texts produced in multiple communicative tanks and accompanied by investories of personality of their respective authors. The corpus is, to the best of our knowledge. the largest resource of this kind available for the Brazilius Portuguene language, and it intends to provide support for a wide range of NLP studies based on personality traits. The rest of this paper is structured as follows. After a brief background discussion (Section 2), the work focuses on the corrun collection task (Section 3) and on its various components (Section 6). Preliminary results of personality recognition from the corpus text are presented for illustration purposes (Section S). This is followed by a discussion on possible applications and extensions of the present work

#### 2. Background

The Big Five model (Coldberg, 1990) comprises five fundamental dimensions of the human personality - Openness to experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism - that may be estimated by using a wide range of methods, the most common being the use of personality inventories. Among many inventories developed for the flig Five model, the need for a fast assessment tool led to the proposal of the 8F7 inventory (John et al.,

The BFI inventory has been replicated in dozens of other languages, including some studies dedicated to our target language, Brazilian Portuguese. In particular, the study in 66c Andrude, 2008) validated the BFI for Brazilian Porturnese through factorial analysis of a sample of 5.089 respondents from all regions of the country. The investory considered in 4de Andrade, 2008) will be the basis of the present work as well.

	April 10 pr	Mark State 11	in terms spired	Bridge Co.	Di anni tiera il	Quality in	
•	300						
		Material 1	MONTH, MINISTRA				
			ajudendo poneja a duri	A seri arross a terre a s	us finantia que sité	OR SHARKS	v SM/MREKS prietry good variou VOTAK NES
			value share since so re	Samer stou share man	uma sea deputs	-	presença de um animal FABULOSO no minor
	100		month parter , A , \$40	HERE MIC SMURRE	RE CLASSICS CAN	P TOUCH THE	S SNUMBERS - Super SNAMES Irreduc SNAME
			SHUMBERS jerons de die	entry entitle if his book	nan pra waste . **	desculps stre	palhar mas vocits podem gaster ENUMBERS :
			matter se wire first fives	de co-autoria de SA	AMES THAMES IN	AMES ENAME	IS SHAMES ENAMES ENAMES SHAMES FROM
	100			7 7 d'aum poqui ale stelle	indana socia manie	a wie shede	so much sringe . I feel said . VOM MONSTIN
			" Well , you resumin a d	pera e espero que visa	Re Secure - Period to	dos meus mir	mence por ser uma pessoa juventi . Aquetes s
			externo That was good	Year I wifer real bro	7 SHAMES . Poles	mons . Allegris	a . Eu não gosto de gatos e sim de pokemon
			som breducação SNU	MEERS SPAINNERS SHE	MINERS exturiance	de use se par	use per enecado no ENAMES , short gritore
			110101010101011	M MATTER STREET	TTTTTTT ENGON	THE OW NAC	ICS ( ) reporturing are 177777777777777
			pi li Lansinja Mecdinocu ja	It Harry Potter E As Re	disputes Dia Morte j	A It Hearny Post	iar E O Prisioneiro De BNAMES ja il Harry Poli
	1004		Chaga de aportado I A fo	otia Brissifetra music bo	rite the cithede with	tou prix mouth	ar que existe amor em 57 . No dia primeiro s
	1000		in perdulo detas confun	tido esta coração . Na	da prode o chido c	-	santido apelo do Não . Ao curios tampinais to
	1000		, mas mão delixou de ser	kndo Yam s	per author sper with	d impulmentive	f . Que vão se fazer a corte e so curtes munos
-	-	14 of 11,537	entries, 3 total columns				

## Datasets documents

**B5 corpus** (public dataset of facebook posts + responses to BFI)

N = 1,082 participants e 194,382 posts e 2,219,585 words (tokens)

We divided posts into chunks of approximately 250 words (11,537) texts segments)

415 items from: SENNA, BFI-2 and OECD Study of SEMS (domains and facets)

	4 4	- 6								
1.9	edit, index/edi	soften.	Hern, pt, best	Non, et, but	domain	Smith	polite	perhan,	pers	
AS.	1,950	362,47	Priso ser militarente, No a distante dos cumos	Can be said and uncerng.	Α	one				
4.1	2 870	142,17	Proce me importe com se cutros	Feels little sympathy for others.	A	Limite.				
4	3 9/9	ME. 94	Tanto competable, um coração mois	to companionate, has a soft heart.	9.	one				
жī.	4 970	160,33	Me preocupo com se suffice a posto de apuda los:	to harpful and uneartish with others.	A	uny	1		2	
4	5 8FU	1442_52	Now educates a gardi com as passess.	As polities, occurrences to other to.	A	798				
ΑE	6.902	MAR. MR.	Contume team on process core requells	is respectful, trade others with respect.	A	799				
4.1	7 870	M42, 57	As veces, you may adjugate som us sumos.	in surregimes made in without.		700			3	
4	8 BFG	142,22	Contume contenter as personal	Starts arguments with others.		-91			4	
88	8 892	MAE 517	Acres (in the bordlade day personal	Assumes the best about people.		TOTAL .				
m E	10 BFG	142,276	Santo a territoria se partinal se presone.	Mass a forgiving mature.		met				
10	11 852	1012 AZ	Conforms duvolar day "bogo intemples" day surios.	to transpolitors of others' propertiess.		and.				
m.	G MPG	642,12	Contacting author districts that making	Tarolic by English and Administration		Page 1				
-	13 6502	ME2.10	Son organizato, gosto de marter an sonas em ortom	to symmetry, New to know things in come:	6	mg	1			
-	14 850	147 ES	Martenhe as stokes imper e ammadas	Names Rivings report and Note	0	100				
	18 8530	142.49	Dates as colors baguings has seen brook.	Leaves a trace, dozen't deservice.		100				
91	16 800	142.3	Coolume ser desergenzado	Tonds to be discountised.		***				
	17 974	942 23	Santo difficultable para company as brodies	The officially parting excited or taking		pred	1			
	18 800	NA2 28	Face on tendra ben's contributed this de-tenue	to efficient, gare things down.	6	prod			**	
	10 870	1642 63	No perceiore, no tedico in tendro all que entigen terrenalis.	in providing works unlittle best in Brighan	0	great				
PI C	20 MFS2	1462 8	Contame car progulpton	Seek triange.		priori			100	
AL	21 BFQ	140,719	She are beds decarateds over many compromisess.	Cast be accrareful parelyss.	- 6	TM				
24	EX MAG	160,590	Sea mas decisionals, describely nations de facul se colors	Cars by somewhat carefuse.		The .			161	
10	23.860	162,13	Toru responsibled, Supr. monte parle de trabalho.	to departure exacts	0	ma.	- 4		11	
n) I	34 850	1662: 7th	New cost destruction below to discuss to	is reflective, care always for counted on	6	766			10	
n.	.25 8FD	162,58	As verse, comports-me de famile insegnmente	Sometimes behaves imagionality.	6	760			91	
201	36 6FQ	142,06	Auto affeit orfuenciar as passesas	Finals inhard to influence people.		med.			12	
		-1	Fain solve morte opmilie, means que soja obresse das							
20.	21 BPG	142.4	Indian promose.	Has an assertine participation.		-	1 1		14	
100	24 850	142.21	Assume contributionals a posição de interesça, de poder	is duminant, acts on a leader.		-	-			
-	29 6912	ME 31	Prefix que os quina tomen so decisões.	Prefers to have others take charge.	- 1	-			14	
-	30 800	142.41	Seu chair de energia	is but of anargy		ang	- 1		10.	
44	21 WILL	142,24	Sou menos plos de que cuitas pessoas-	to been earliest than other persons.		***			**	
П	-		Ranamenta ainto-ma sangolganto no com multas asportativos							
	N Mrs	142.11	sites size.	Plants have excluded an excess.		-		6 0	-	
-	16 97 9		Total Addition	The state of the s	_	-	_		-	

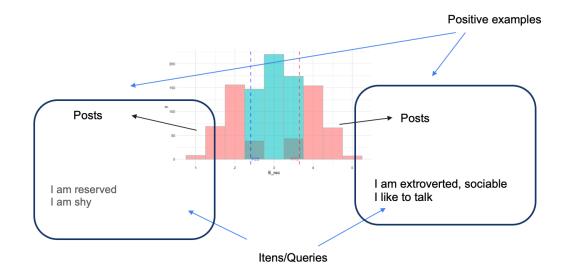
## Dataset creation for finetuning BERT

I paired posts with items

415 items X 11537 texts = 4.787.855 possible interactions

Marked positive examples

Marked negative example



Then I created triplets:

post vs positive item example vs negative item example

#### Nils Reimers and Iryna Gurevych

Ubiquitous Knowledge Processing Lab (UKP-TUDA)
Department of Computer Science, Technische Universität Darmstadt
www.ukp.tu-darmstadt.de

#### Abstract

BERT (Devin et al., 2018) and RoBERTa (Liu et al., 2019) has set a new state-of-the-art performance on sentence-pair regression tasks like semantic textual similarity (STS). However, it requires that both sentences are fed into the network, which causes a massive computational overhead: Finding the most similar pair in a collection of 10,000 sentences requires about 50 million inference computations (-65 hours) with BERT. The construction of BERT makes it unsuitable for semantic similarity search as well as for unsupervised tasks like clustering.

In this publication, we present Sentence-BERT (SBERT), a modification of the pretrained BERT network that use siamese and triplet network structures to derive semantically meaningful sentence embeddings that can be compared using cosine-similarity. This reduces the effort for finding the most similar pair from 65 hours with BERT / RoBERTa to about 5 seconds with SBERT, while maintaining the accuracy from BERT.

We evaluate SBERT and SRoBERTa on common STS tasks and transfer learning tasks, where it outperforms other state-of-the-art sentence embeddings methods.<sup>1</sup>

#### 1 Introduction

In this publication, we present Sentence-BERT (SBERT), a modification of the BERT network using siamese and triplet networks that is able to derive semantically meaningful sentence embeddings<sup>2</sup>. This enables BERT to be used for certain new tasks, which up to now were not applicable for BERT. These tasks include large-scale seman-

tic similarity comparison, clustering, and information retrieval via semantic search.

BERT set new state-of-the-art performance on various sentence classification and sentence-pair regression tasks. BERT uses a cross-encoder: Two sentences are passed to the transformer network and the target value is predicted. However, this setup is unsuitable for various pair regression tasks due to too many possible combinations. Finding in a collection of n = 10000 sentences the pair with the highest similarity requires with BERT  $n \cdot (n-1)/2 = 49\,995\,000$  inference computations. On a modern V100 GPU, this requires about 65 hours. Similar, finding which of the over 40 million existent questions of Quora is the most similar for a new question could be modeled as a pair-wise comparison with BERT, however, answering a single query would require over 50 hours.

A common method to address clustering and semantic search is to map each sentence to a vector space such that semantically similar sentences are close. Researchers have started to input individual sentences into BERT and to derive fixedsize sentence embeddings. The most commonly used approach is to average the BERT output layer (known as BERT embeddings) or by using the output of the first token (fte [CLS] token). As we will show, this common practice yields rather bad sentence embeddings, often worse than averaging GloVe embeddings (Pennington et al., 2014).

To alleviate this issue, we developed SBERT. The siamese network architecture enables three-direct evectors for input sentences can be derived. Using a similarity measure like cosine-similarity or Manhatten / Buclidean distance, semantically similar sentences can be found. These similarity measures can be performed extremely efficient on modern hardware, allowing SBERT to be used for semantic similarity search as well as for clustering. The complexity for finding the

#### ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT

Ornar Khattab Stanford University okhattab@stanford.edu

#### ABSTRACT

2020

:2004

Recent progress in Natural Language Understanding (NLU) is driving fast-paced advances in Information Retrieval (IR), largely owed to fine-tuning deep language models (LMs) for document ranking. While remarkably effective, the ranking models based on these LMs increase computational cost by orders of magnitude over prior approaches, particularly as they must feed each query-document pair through a massive neural network to compute a single relevance score. To tackle this, we present ColBERT, a novel ranking model that adapts deep LMs (in particular, BERT) for efficient retrieval. ColBERT introduces a late interaction architecture that independently encodes the query and the document using BERT and then employs a cheap yet powerful interaction step that models their fine-grained similarity. By delaying and yet retaining this finegranular interaction, ColBERT can leverage the expressiveness of deep LMs while simultaneously gaining the ability to pre-compute document representations offline, considerably speeding up query processing. Beyond reducing the cost of re-ranking the documents retrieved by a traditional model. ColBERT's pruning friendly in teraction mechanism enables leveraging vector-similarity indexes for end-to-end retrieval directly from a large document collection. We extensively evaluate ColBERT using two recent passage search datasets. Results show that ColBERT's effectiveness is competitive with existing BERT-based models (and outperforms every non BERT baseline), while executing two orders-of-magnitude faster and requiring four orders-of-magnitude fewer FLOPs per query.

#### ACM Reference form

Omar Khattab and Matez Zaharis. Não. ColBERT Efficient and Effective Passage Search via Contestualizate Liae Interaction over IEET. In Proceedings of Proceedings of the 45rd International ACM NSGR Conference on Research and Developments in Information Retrieval, Virtual Event, China, July 25–30, 2020 (SASR '20). 10 pages. DOI: 10.1145/3977.3-3401075

#### 1 INTRODUCTION

Over the past few years, the Information Retrieval (IR) community has witnessed the introduction of a host of neural ranking models, including DRMM [7], KNRM [4, 36], and Duet [20, 22]. In contrast

DOE 10.1145/3397271.3461075

Matei Zaharia Stanford University matei@cs.stanford.edu

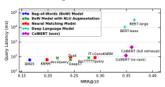


Figure 1: Effectiveness (MRR@10) versus Mean Query Latency (log-scale) for a number of representative ranking models on MS MARCO Ranking [24]. The figure also shows ColBERT. Neural re-rankers run on top of the official BM25 top-1000 results and use a Tesla V100 GPU. Methodology and detailed results are in §4.

to prior learning-to-rank methods that rely on hand-crafted features, these models employ embedding-based representations of queries and documents and directly model local interactions (i.e., fine granular relationships) between their contents. Among them, a recent approach has emerged that four name deep per trained language models (LMs) like ELMs [29] and BERT [5] for estimating relevance. By competing deeply-contestualized semantic representations of query-document pairs, these LMs help bridge the pervasive vocabulary mismatch [21, 42] between documents and queries [30]. Indeed, in the span of just a few months, a number of ranking models based on BERT have achieved state of the art results on various retrieval benchmarks [5, 18, 28, 39] and have been proprietarly adapted for deproyment by Coople and Bing<sup>2</sup>.

However, the remarkable gains delivered by these LMs come at a steep increase in computational cost Hofsitäte et al. [9] and MacAvancy et al. [18] observe that BRET based models in the literature are 100 1000× more computationally expensive than prior models—some of which are arguably not incapenave to begin with [13]. This quality—cost tradeoff is summarized by Figure 1, which compares two BRET based randers [28, 27] against a representative set of ranking models. The figure uses MS MARCO Ranking [24], a recent collection of 9M passages and 1M queries from Bing's a recent collection of 9M passages and 1M queries from Bing's alight-ind server that dedicates one Tesla V100 GPU per query for neural re-rankers. Following the \*re-ranking\* servey of MS MARCO. Collibert (re-rank), the Neural Matching Models, and the Deep LMs er and the Server and the MS MARCO's official top 1000 documents per query.

<sup>&</sup>lt;sup>1</sup>Code available: https://github.com/UKPLab/ sentence-transformers

<sup>&</sup>lt;sup>2</sup>With semantically meaningful we mean that semantically similar sentences are close in vector space.

Permassive to make digital or hard copies of all or part of this words for presented or clustersom uses a general without fee provided that copies a hear than seed a distributed fee profit or commercial advantage and that copies here this notice and the full critiston on the first page. Corputalis face components of this work works soly whether than the cushwist must be honored. Advantaging with credit is permitted to super-observior, or republish, to post on severes a time distribute in lest, requires pies up on the permission and/or a fee. Request permissions from permissions/blant org. SSGP 30, Person (Pross, Chine

<sup>© 2026</sup> Copyright held by the owner/author(s). Publication rights licensed to ACM. 978-1-4585-8016-4/28/07... \$15.00

https://blog.google/products/search/search-language-understanding-bert/ "https://autre.microsoft.com/en-ua/blog/bing-delivers-its-largest-improvementin-search-experience-unine-unine-unine-uni-

## Sentence Transformer

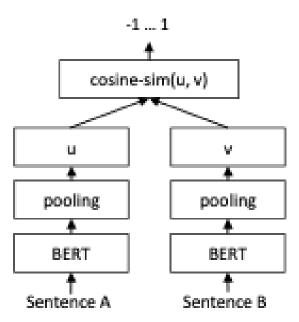


Figure 2: SBERT architecture at inference, for example, to compute similarity scores. This architecture is also used with the regression objective function.

## ColBERT

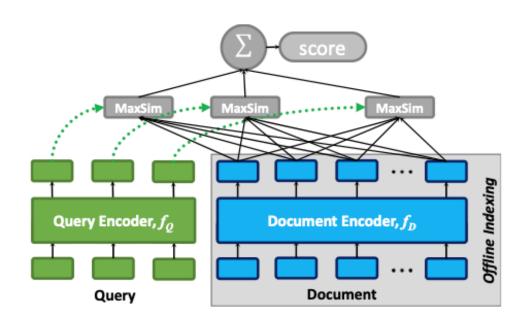
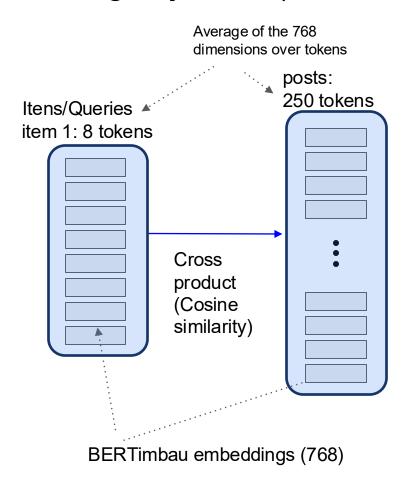


Figure 3: The general architecture of ColBERT given a query q and a document d.

## Training objective (loss function)



#### **Training Objective**

Similarity between positive examples is greather than negaive examples

#### **Training and Testing Data:**

Training Set: 993,759 interactions (N=868) Test Set: 158,594 interactions (N=152)

#### **Training Details:**

Epochs: 3 (each epoch took 6 hours and 40 minutes in a GPU Nvidia A100 batches of 80 triplets)

The model used is BERTimbau, pretrained in Portuguese (https://github.com/neuralmind-ai/portuguese-bert).

#### Model Application (scoring posts for Big Five):

Used the model to score each interaction item by post in the test set (Created semantic similarity scores: sim\_scores for each item int he item database)

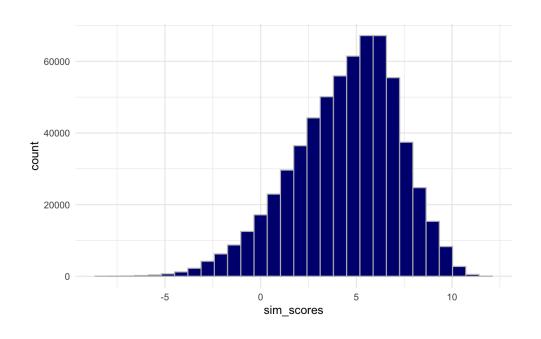
**Generated scores for domains** and facets by averaging sim\_scores for each domain/facet

post vs itens of domain

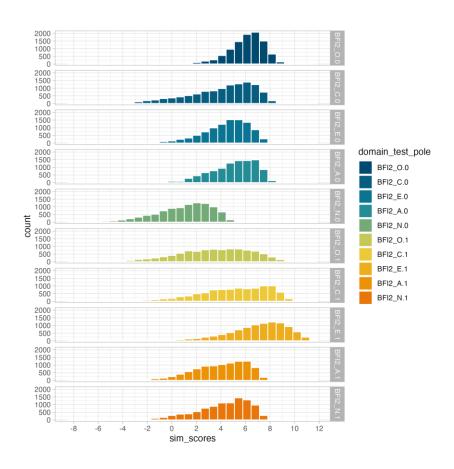
## **Findings**

- What most common theme of big five was seen in posts?
- What is the relationship between subject BFI scores and personality scores from their posts?

## What is the distribution of sim\_scores?



### What is the distribution of sim\_scores by domain / pole test of B5?

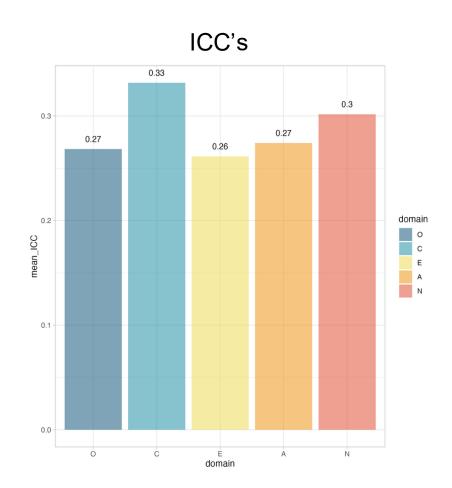


### Multilevel structure of the data

Person (Level 2)

Persons + domain

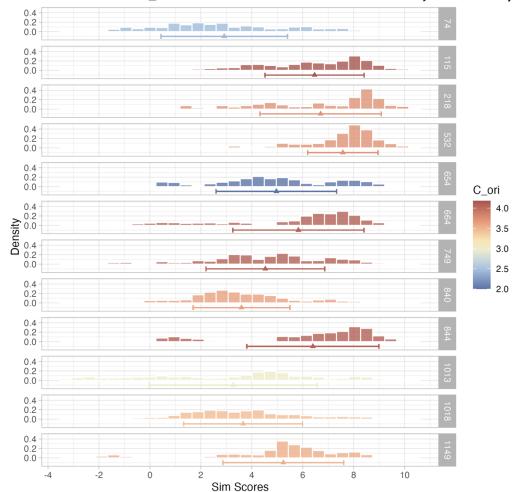
FB Posts (Level 1)



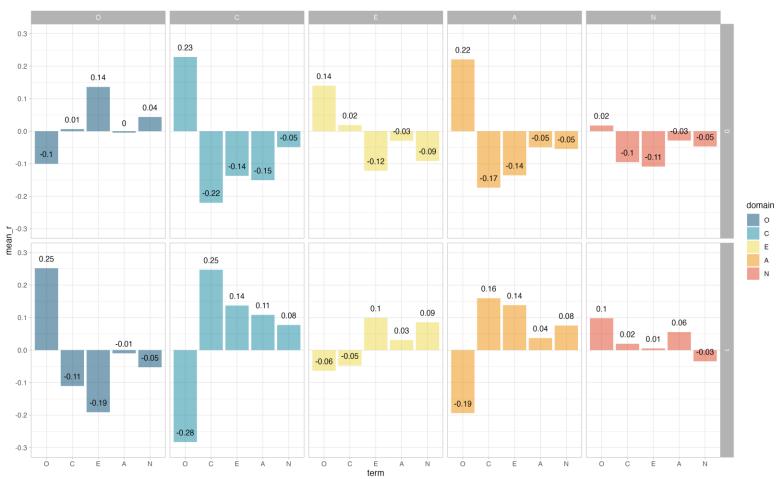
### **Probabilistic Density Function**

Distributions of sim\_scores for C of 12 randomly selected subjects

Distribution of sim\_scores on Conscientiousness for 12 randomly selected subje



## How self-reported B5 (BFI) correlates with sim\_scores by domain and pole?



## Regression predicting self-reported B5 in BFI from average sim scores of test/domain/pole

Journal of Personality and Scool Prostolog-

0.2014 American Psychologosii Association 0022-3514/14512.00 http://dx.doi.org/10.3017[psyp0000020

#### Automatic Personality Assessment Through Social Media Language

Gregory Park, H. Andrew Schwartz, Johannes C. Eichstaedt, and Margaret L. Kern University of Pennsylvania

Michal Kosinski and David J. Stillwell University of Cambridge

Lyle H. Ungar and Martin E. P. Seligman University of Pennsylvania

Language use is a psychologically rich, stable individual difference with well-established correlations to personality. We describe a method for assessing personality using an open-wocabulary analysis of language from social media. We compiled the written language from 66,732 Pacebook users and their questionnaire-based self-reported Big Five personality traits, and then we built a predictive model of personality based on their language. We used this model to predict the 5 personality factors in a separate sample of 4.824 Facebook users, examining (a) convergence with self-reports of personality at the domain- and facet-level: (b) discriminant validity between predictions of distinct traits: (c) agreement with informant reports of personality; (d) patterns of correlations with external criteria (e.g., number of friends, political attitudes, impulsiveness); and (e) test retest reliability over 6-month intervals, Results indicated that language-based assessments can constitute valid personality measures: they agreed with self-reports and informant reports of personality, added incremental validity over informant reports, adequately discriminated between traits, exhibited natterns of correlations with external criteria similar to those found with self-reported personality, and were stable over 6-month intervals. Analysis of predictive language can provide rich portraits of the mental life associated with traits. This approach can complement and extend traditional methods, necoding researchers with an additional measure that can quickly and cheaply assess large groups of participants with minimal burden.

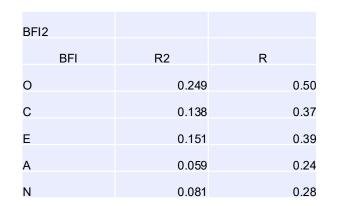
Keywords: language, personality assessment, measurement, big data, social media

Supplemental materials: http://dx.doi.org/10.1037/pspp0000020.supp

Table 2 Correlations Between Language-Based Assessments and Self-Reports of Big Five Personality Traits

			elf-report	bs	L	Language-based assessments							
	0	C	E	A	N	0	C	E	A	N			
Self-reports													
Openness													
Conscientiousness	.00												
Extraversion	.13	.19											
Agreeableness	.07	.17	.19										
Neuroticism	08	- 31	34	36									
Language-based													
Openness	.43	12	-0.08	05	.00.								
Conscientiousness	13	.37	.16	.17	17	25							
Extraversion	07	.12	.42	.10	15	17	33						
Agreeableness	07	.17	.13	.35	14	12	.44	.27					
Neuroticism	.05	17	18	13	.35	.06	=.41	43	- 34				

Note: N = 4.824, O = Openness to Experience: C = Conscientiousness: E = Extraversion: A = Agreeableness: N = Neuroticism. Convergent correlations are in bold: discriminant correlations are in italics.



## Conclusions and discussion

- Benchmarking Results:
  - Our methods consistently achieved the upper-level benchmarks.
     Non trait-activated
- Innovative Methodology
  - Step 1 Assessed texts scoring the big five. Step 2. Aggregated by domain to score subjects
  - O General Text Analysis by psychological themes: Provides a new, effective way to study texts, harnessing the richness of spontaneous, user-generated content.
  - Automated Insights: Leverages AI to gain psychological insights
  - Scalability: The method is adaptable and can be applied to any test!
- Bridging AI and Psychology:
  - Al Enhancing Psychology: Utilizes advanced Al techniques to deepen psychological understanding, offering more nuanced and precise analyses of human behavior.
  - Psychology Informing AI: Psychological theories guide the development of AI models, ensuring they are more attuned to human nuances and behaviors.

## Resources

Behavior Research Methods https://doi.org/10.3758/s13428-024-02455-8

ORIGINAL MANUSCRIPT



### A tutorial on open-source large language models for behavioral science

Zak Hussain 1,2 - Marcel Binz 3,4 - Rui Mata 1 - Dirk U. Wulff 1,2

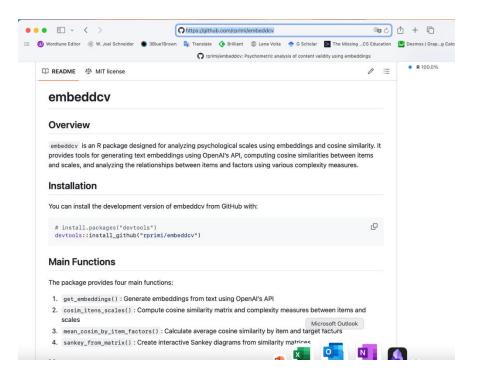
Accepted: 27 May 2024 © The Author(s) 2024

#### Abstract

Large language models (LLMs) have the potential to revolutionize behavioral science by accelerating and improving the research cycle, from conceptualization to data analysis. Unlike closed-source solutions, open-source frameworks for LLMs can enable transparency, reproducibility, and adherence to data protection standards, which gives them a crucial advantage for use in behavioral science. To help researchers harmess the promise of LLMs, this tutorial offers a primer on the open-source Hugging Face ecosystem and demonstrates several applications that advance conceptual and empirical work in behavioral science, including feature extraction, fine-tuning of models for prediction, and generation of behavioral resones. Executable code is made available at github.com/Zak-Hussain/LLM4BeSci.git. Finally, the tutorial discusses challenges faced by research with (open-source) LLMs related to interpretability and safety and offers a perspective on future research at the intersection of language modeling and behavioral science.

Keywords Large language models - Behavioral science - Hugging face

### https://github.com/rprimi/embeddcv

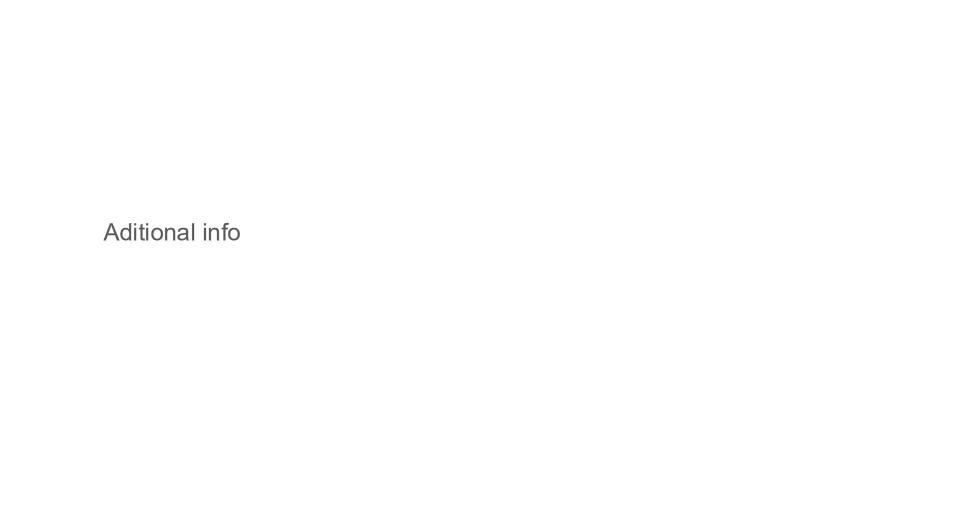


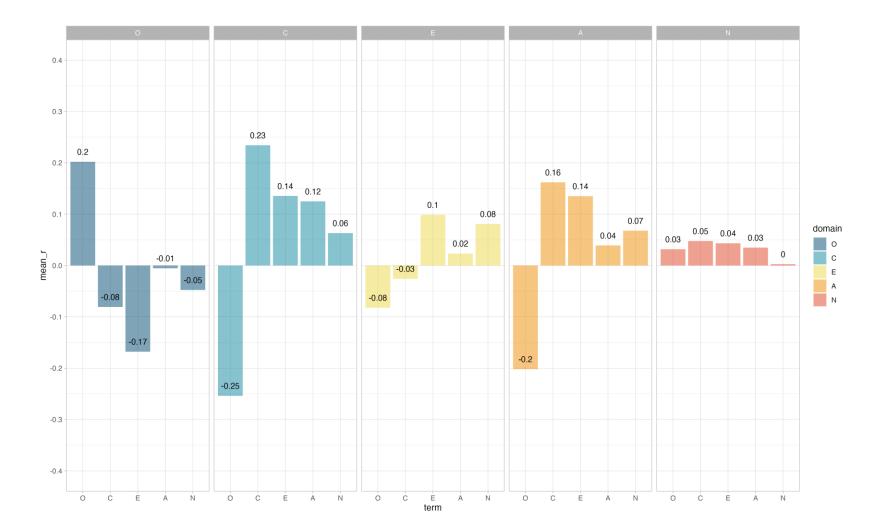
## Thank You ... rprimi@mac.com











vui s	0_100			7-1-00	
BFI2_A_cmp_0	0.15	-0.14	-0.08	-0.03	-0.03
2_A_cmp_1	-0.13	0.12	0.06	0.03	0.06
2_A_res_0	0.16	-0.14	-0.08	-0.03	-0.04
SFI2_A_res_1	-0.13	0.12	0.06	0.03	0.05
BFI2_A_trst_0	0.16	-0.14	-0.08	-0.03	-0.04
BFI2 A trst 1	-0.13	0.12	0.06	0.03	0.06
BFI2 C org 0	0.18	-0.16	-0.14	-0.10	-0.04
BFI2_C_org_1	-0.22	0.19	0.07	0.08	0.06
BFI2_C_prod_0	0.18	-0.16	-0.14	-0.10	-0.04
BFI2_C_prod_1	-0.22	0.19	0.07	0.08	0.06
BFI2_C_prod_1	0.19	-0.16	-0.14	-0.10	-0.04
BFI2_C_rsp_1	-0.22	0.19	0.07	0.08	0.06
BFI2_E_ass_0	0.12	0.19	-0.18	-0.04	-0.06
	-0.06	-0.04	0.14	0.04	0.04
BFI2_E_ass_1					
BFI2_E_enrg_0	0.12	0.01	-0.18	-0.04	-0.07
BFI2_E_enrg_1	-0.06	-0.05	0.14	0.04	0.04
BFI2_E_soc_0	0.12	0.01	-0.18	-0.04	-0.06
BFI2_E_soc_1	-0.06	-0.05	0.14	0.04	0.04
BFI2_N_anx_0	0.07	-0.07	-0.11	-0.03	-0.04
BFI2_N_anx_1	0.06	0.02	0.02	0.04	0.01
BFI2_N_dep_0	0.07	-0.07	-0.11	-0.03	-0.04
BFI2_N_dep_1	0.06	0.02	0.01	0.04	0.01
BFI2_N_vlti_0	0.07	-0.07	-0.11	-0.03	-0.04
BFI2_N_vlti_1	0.06	0.02	0.01	0.04	0.01
BFI2_O_aes_0	-0.11	0.06	0.04	-0.01	0.02
BFI2_O_aes_1	0.22	-0.12	-0.10	0.01	-0.03
BFI2_O_crea_0	-0.11	0.06	0.04	-0.01	0.02
BFI2_O_crea_1	0.22	-0.12	-0.10	0.01	-0.03
BFI2_O_int_0	-0.11	0.05	0.04	-0.01	0.02
BFI2_O_int_1	0.22	-0.12	-0.10	0.01	-0.03

vars

O\_rec

C\_rec

E\_rec

A\_rec

N\_rec

O\_rec C\_rec E\_rec A\_rec N\_rec

vars

vars	O_rec	C_rec	E_rec	A_rec	N_rec
SENNA_A_Cmp_0	0.16	-0.14	-0.08	-0.03	-0.04
SENNA_A_Cmp_1	-0.13	0.12	0.06	0.03	0.05
SENNA_A_Mod_0	0.15	-0.14	-0.08	-0.03	-0.03
SENNA_A_Mod_1	-0.13	0.12	0.06	0.03	0.06
SENNA A Resp 0	0.15	-0.14	-0.07	-0.04	-0.03
SENNA_A_Resp_1	-0.13	0.12	0.06	0.03	0.05
SENNA_A_Tru_0	0.16	-0.14	-0.08	-0.04	-0.04
SENNA_A_Tru_1	-0.13	0.12	0.07	0.03	0.06
SENNA_C_Achv_0	0.18	-0.16	-0.13	-0.10	-0.04
SENNA_C_Achv_1	-0.22	0.19	0.07	0.08	0.06
SENNA_C_Conc_0	0.18	-0.16	-0.14	-0.10	-0.04
SENNA_C_Conc_1	-0.22	0.19	0.07	0.08	0.06
SENNA_C_Ord_0	0.18	-0.16	-0.14	-0.10	-0.04
SENNA_C_Ord_1	-0.22	0.19	0.07	0.08	0.06
SENNA_C_SD_0	0.19	-0.16	-0.14	-0.10	-0.04
SENNA_C_SD_1	-0.22	0.19	0.07	0.08	0.06
SENNA_C_SofR_0	0.19	-0.16	-0.14	-0.10	-0.04
SENNA_C_SofR_1	-0.22	0.19	0.07	0.08	0.06
SENNA E Act 0	0.12	0.01	-0.18	-0.04	-0.07
SENNA E Act 1	-0.06	-0.05	0.14	0.04	0.04
SENNA E Assr 0	0.12	0.01	-0.18	-0.04	-0.07
SENNA_E_Assr_1	-0.07	-0.04	0.15	0.04	0.04
SENNA_E_Soc_0	0.12	0.01	-0.18	-0.04	-0.07
SENNA E Soc 1	-0.06	-0.04	0.15	0.04	0.04
SENNA_N_LAngrVol_0	0.07	-0.07	-0.11	-0.03	-0.04
SENNA_N_LAngrVol_1	0.07	0.02	0.01	0.04	0.01
SENNA_N_LAnx_0	0.07	-0.07	-0.11	-0.03	-0.04
SENNA_N_LAnx_1	0.06	0.02	0.01	0.04	0.01
SENNA N LDep 0	0.07	-0.07	-0.11	-0.03	-0.04
SENNA_N_LDep_1	0.06	0.02	0.01	0.04	0.01
SENNA_O_Aes_0	-0.10	0.05	0.04	-0.01	0.01
SENNA O Aes 1	0.22	-0.12	-0.10	0.01	-0.03
SENNA_O_CrImg_0	-0.11	0.06	0.04	-0.01	0.02
SENNA_O_CrImg_1	0.22		-0.10	0.01	-0.03
SENNA_O_IntCur_0	-0.10	0.05	0.04	-0.01	0.01
SENNA_O_IntCur_1	0.22	-0.12	-0.10	0.01	-0.03
JEHNA_O_IIICGI_I	0.22	-0.12	-0.10	0.01	-0.03

BFI2_A_cmp_0	0.26	-0.23	0.00	-0.09	0.01	OCDE_SEMS_A_COO_0	0.25	-0.23	0.00	-0.09	0.01	SENNA_A_Cmp_0	0.26	-0.23	0.00	-0.09	0.01
BFI2_A_cmp_1	-0.22	0.21	-0.07	0.08	0.04	OCDE_SEMS_A_COO_1	-0.22	0.22	-0.07	0.08	0.04	SENNA_A_Cmp_1	-0.21	0.21	-0.07	0.08	0.04
BFI2 A res 0	0.26	-0.23	0.00	-0.09	0.01	OCDE_SEMS_A_EMP_0	0.27	-0.24	0.00	-0.09	0.01	SENNA_A_Mod_0	0.26	-0.23	0.00	-0.09	0.01
BFI2_A_res_1	-0.21	0.21	-0.07	0.08	0.04	OCDE_SEMS_A_EMP_1	-0.21	0.21	-0.07	0.08	0.04	SENNA_A_Mod_1	-0.22	0.22	-0.07	0.08	0.04
BFI2 A trst 0	0.26	-0.24	0.00	-0.09	0.01	OCDE_SEMS_A_TRU_0	0.26	-0.23	0.00	-0.09	0.01	SENNA_A_Resp_0 SENNA_A_Resp_1	0.25 -0.21	-0.23 0.21	0.00 -0.07	-0.09 0.08	0.01
BFI2_A_trst_1	-0.22	0.22	-0.07	0.08	0.04	OCDE_SEMS_A_TRU_1	-0.22	0.22	-0.07	0.08	0.04	SENNA_A_Resp_1 SENNA A Tru 0	0.21	-0.24	0.00	-0.09	0.04
BFI2_C_org_0	0.20	-0.21	-0.13	-0.12	-0.08	OCDE_SEMS_C_MOT_0	0.20	-0.21	-0.13	-0.12	-0.08	SENNA A Tru 1	-0.22	0.22	-0.07	0.08	0.01
						OCDE_SEMS_C_MOT_1	-0.34	0.27	0.02	0.11	0.09	SENNA C Achy 0	0.20	-0.21	-0.13	-0.12	-0.08
BFI2_C_org_1	-0.34	0.27	0.02	0.11	0.09	OCDE_SEMS_C_PER_0	0.20	-0.21	-0.13	-0.12	-0.08	SENNA_C_Achv_1	-0.34	0.27	0.02	0.11	0.09
BFI2_C_prod_0	0.20	-0.21	-0.13	-0.12	-0.08	OCDE_SEMS_C_PER_1	-0.34	0.27	0.02	0.11	0.09	SENNA_C_Conc_0	0.20	-0.21	-0.13	-0.12	-0.08
BFI2_C_prod_1	-0.34	0.27	0.02	0.11	0.10	OCDE_SEMS_C_RES_0	0.20	-0.21	-0.13	-0.12	-0.08	SENNA_C_Conc_1	-0.34	0.27	0.02	0.11	0.09
BFI2_C_rsp_0	0.20	-0.21	-0.13	-0.12	-0.08	OCDE_SEMS_C_RES_1	-0.34	0.27	0.02	0.11	0.09	SENNA_C_Ord_0	0.20	-0.21	-0.13	-0.12	-0.08
BFI2_C_rsp_1	-0.34	0.27	0.02	0.11	0.09	OCDE_SEMS_C_SEL_0	0.20	-0.21	-0.13	-0.12	-0.08	SENNA_C_Ord_1	-0.34	0.27	0.02	0.11	0.09
BFI2_E_ass_0	0.09	-0.10	-0.24	-0.20	-0.03	OCDE_SEMS_C_SEL_1	-0.34	0.27	0.02	0.11	0.09	SENNA_C_SD_0	0.20	-0.21	-0.13	-0.12	-0.08
BFI2_E_ass_1	-0.01	0.05	0.25	0.20	-0.02	OCDE_SEMS_E_ASS_0	0.09	-0.10	-0.25	-0.20	-0.03	SENNA_C_SD_1	-0.34	0.27	0.02	0.11	0.09
BFI2_E_enrg_0	0.09	-0.10	-0.25	-0.20	-0.04	OCDE_SEMS_E_ASS_1	-0.01	0.05	0.25	0.20	-0.02	SENNA_C_SofR_0	0.20	-0.21	-0.13	-0.12	-0.08
BFI2 E enrg 1	-0.01	0.04	0.25	0.20	-0.03	OCDE_SEMS_E_ENE_0	0.09	-0.10	-0.25	-0.20	-0.04	SENNA_C_SofR_1	-0.34	0.27	0.02	0.11	0.09
BFI2 E soc 0	0.08	-0.10	-0.25	-0.20	-0.04	OCDE_SEMS_E_ENE_1	-0.01	0.05	0.25	0.20	-0.02	SENNA_E_Act_0	0.09	-0.10	-0.24	-0.20	-0.04 -0.03
BFI2_E_soc_1	-0.01	0.04	0.25	0.20	-0.03	OCDE_SEMS_E_SOC_0	0.09	-0.10	-0.25	-0.20	-0.03	SENNA_E_Act_1 SENNA E Assr 0	-0.01 0.10	0.04 -0.11	0.25 -0.24	0.20 -0.20	-0.03
BFI2_N_anx_0	0.03	-0.08	-0.18	0.00	-0.07	OCDE_SEMS_E_SOC_1	-0.01	0.05	0.25	0.20	-0.03	SENNA_E_ASSI_0 SENNA_E_ASSI_1	-0.01	0.05	0.25	0.20	-0.03
BFI2_N_anx_1	0.18	-0.07	0.08	-0.02	0.12	OCDE_SEMS_N_EMO_0	0.03	-0.08	-0.18	0.00	-0.08	SENNA_E_Soc_0	0.09	-0.10	-0.24	-0.20	-0.04
BFI2 N dep 0	0.04	-0.08	-0.18	0.00	-0.08	OCDE_SEMS_N_EMO_1	0.19	-0.07	0.08	-0.02	0.12	SENNA_E_Soc_1	-0.01	0.05	0.25	0.20	-0.02
		-0.07	0.08	-0.02		OCDE_SEMS_N_OPT_0	0.03	-0.08	-0.18	0.00	-0.08	SENNA_N_LAngrVol	0.04	-0.08	-0.18	0.00	-0.08
BFI2_N_dep_1	0.19				0.12	OCDE_SEMS_N_OPT_1	0.18	-0.07	0.08	-0.02	0.12	SENNA_N_LAngrVol	0.19	-0.08	0.07	-0.02	0.11
BFI2_N_vlti_0	0.03	-0.08	-0.18	0.00	-0.08	OCDE_SEMS_N_STR_0	0.03	-0.08	-0.18	0.00	-0.08	SENNA_N_LAnx_0	0.03	-0.08	-0.18	0.00	-0.08
BFI2_N_vlti_1	0.19	-0.07	0.08	-0.02	0.12	OCDE_SEMS_N_STR_1	0.19	-0.07	0.08	-0.02	0.12	SENNA_N_LAnx_1	0.19	-0.07	0.08	-0.02	0.12
BFI2_O_aes_0	-0.22	0.13	-0.02	-0.06	0.01	OCDE_SEMS_O_CRE_0	-0.20	0.12	-0.03	-0.07	0.01	SENNA_N_LDep_0	0.03	-0.08	-0.18	0.00	-0.08
BFI2_O_aes_1	0.35	-0.18	-0.03	0.07	-0.08	OCDE_SEMS_O_CRE_1	0.35	-0.18	-0.03	0.07	-0.08	SENNA_N_LDep_1	0.19	-0.07	0.08	-0.02	0.12
BFI2_O_crea_0	-0.21	0.12	-0.02	-0.06	0.01	OCDE_SEMS_O_CUR_0	-0.21	0.12	-0.03	-0.06	0.01	SENNA_O_Aes_0	-0.20	0.11	-0.03	-0.07	0.00
BFI2_O_crea_1	0.35	-0.18	-0.03	0.07	-0.08	OCDE_SEMS_O_CUR_1	0.35	-0.18	-0.03	0.07	-0.08	SENNA_O_Aes_1	0.35	-0.18	-0.03	0.07	-0.08
BFI2_O_int_0	-0.21	0.12	-0.02	-0.06	0.01	OCDE_SEMS_O_TOL_0	-0.20	0.12	-0.03	-0.07	0.00	SENNA_O_CrImg_0	-0.21	0.12	-0.03	-0.06	0.01
BFI2_O_int_1	0.35	-0.18	-0.03	0.07	-0.08	OCDE_SEMS_O_TOL_1	0.35	-0.18	-0.03	0.06	-0.08	SENNA_O_CrImg_1 SENNA_O_IntCur_0	0.35 -0.20	-0.18	-0.03 -0.03	0.07 -0.07	-0.08 0.00
												SENNA_O_IntCur_0	0.35	0.11 -0.18	-0.03	0.07	-0.08
												SCHIN_O_IIICUI_I	0.55	-0.10	-0.03	0.07	-0.00

O\_rec C\_rec

E\_rec

A\_rec

N\_rec

O\_rec C\_rec E\_rec

A\_rec

N\_rec

vars

O\_rec C\_rec

E\_rec

A\_rec

N\_rec

## References

Hommel, B. E., Wollang, F.-J. M., Kotova, V., Zacher, H., & Schmukle, S. C. (2022). Transformer-based deep neural language modeling for construct-specific automatic item generation. *Psychometrika*, 87, 749–772. <a href="https://doi.org/10.1007/s11336-021-09823-9">https://doi.org/10.1007/s11336-021-09823-9</a>

Mehta, Y., Majumder, N., Gelbukh, A., & Cambria, E. (2020). Recent Trends in Deep Learning Based Personality Detection. *Artificial Intelligence Review*, 53(4), 2313–2339. <a href="https://doi.org/10.1007/s10462-019-09770-z">https://doi.org/10.1007/s10462-019-09770-z</a>

Plank, B., & Hovy, D. (2015). Personality Traits on Twitter—Or—How to Get 1,500 Personality Tests in a Week. *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 92–98. https://doi.org/10.18653/v1/W15-2913

Santos, W. R. dos, & Paraboni, I. (2019). *Personality facets recognition from text* (arXiv:1810.02980). arXiv. <a href="https://doi.org/10.48550/arXiv.1810.02980">https://doi.org/10.48550/arXiv.1810.02980</a>

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117–143. https://doi.org/10.1037/pspp0000096

Stajner, S., & Yenikent, S. (2020). A Survey of Automatic Personality Detection from Texts. *Proceedings of the 28th International Conference on Computational Linguistics*, 6284–6295. https://doi.org/10.18653/v1/2020.coling-main.553

