



Harmony

Open source NLP/AI tool for psychologists and
social and health sciences
Data discovery and harmonisation

harmonydata.ac.uk
github.com/harmonydata/

Thomas Wood
 **Fast Data Science** fastdatascience.com





MEET THE TEAM



**Eoin
McElroy**

Psychologist
Ulster
University



**Bettina
Moltrecht**

Psychologist
University College
London



**Mauricio
Hoffmann**

Psychologist
Universidade
Federal de Santa
Maria



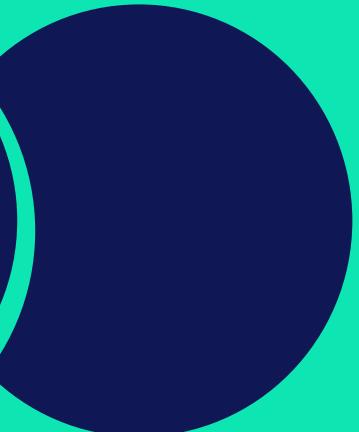
**George
Ploubidis**

Psychologist
University College
London



**Thomas
Wood**

Data science/software
Fast Data
Science



**John
Rogers**

Front end
development



CONTRIBUTORS

**...plus people around
the world who have
made pull requests to
Github!**

**shahid-0
EveWCheng
0x48piraj
nikhildevre
omtarful**

Filters ▾ Q: is:pr is:closed

Clear current search query, filters, and sorts

0 Open ✓ 6 Closed

Add topic strengths ✓
#29 by woodthom2 was merged 2 weeks ago 9 of 12 tasks

Setup GitHub workflow
#27 by shahid-0 was closed 3 weeks ago

Setup pyproject file ✓
#26 by shahid-0 was merged 3 weeks ago 9 of 13 tasks

broke down matcher function into smaller functions ✓
#9 by EveWCheng was merged on Dec 22, 2023

Parserfixesdec2023 ✓
#6 by woodthom2 was merged on Dec 11, 2023

Improving pdf extraction ✘
#1 by woodthom2 was merged on Jun 26, 2023

SECTION



Partners



Economic
and Social
Research Council



GAD-7 Anxiety

Over the <u>last two weeks</u> , how often have you been bothered by the following problems?	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious, or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid, as if something awful might happen	0	1	2	3

Column totals _____ + _____ + _____ + _____ =

Total score _____



Average GAD-7 scores in lockdown vs outside lockdown





GAD-7 Anxiety

Over the last two weeks, have you been bothered by the following symptoms?

1.	<input type="checkbox"/>
2.	<input type="checkbox"/>
3.	<input type="checkbox"/>
4.	<input type="checkbox"/>
5.	<input type="checkbox"/>
6.	<input type="checkbox"/>
7.	<input type="checkbox"/>

Hamilton

Below is a list of symptoms. Please indicate to which degree each symptom applies to you.

0 = Not present

1 = Anxiety

Worries, anxiety

2 = Tension

Feelings of tension, easily, tremble

3 = Fears

Of dark, of strangers, of being left alone, of animals, of traffic, of crowds.

Not difficult

If you can't

things

Pensando nos últimos seis meses, você:

	Nunca	Uma vez	Duas vezes	Três vezes	Quatro vezes	Cinco vezes	Mais de cinco vezes	NÃO SABE
DSHI1. Cortou os pulsos, braços ou outras partes do corpo	0	1	2	3	4	5	6	99
DSHI2. Queimou-se com cigarros, isqueiros ou fósforos	0	1	2	3	4	5	6	99
DSHI3. Fez palavras, ou desenhos na pele através de cortes	0	1	2	3	4	5	6	99
DSHI4. Arranhou-se gravemente a ponto de sangrar	0	1	2	3	4	5	6	99
DSHI5. Mordeu-se a ponto de abrir a pele	0	1	2	3	4	5	6	99

feelings, missing beat.

0 1 2 3 4

10 Respiratory symptoms

0 1 2 3 4

Pressure or constriction in chest, choking feelings, sighing, dyspnea.



Combining studies

Imagine we have response from UK and Brazil

UK: Millennium cohort study
19,000 children

Brazilian High Risk Cohort Study for Childhood Mental Health Conditions (BHRCS).
2,511 young people

Surveys were in different languages, using different instruments

GAD-7 Anxiety

Over the <u>last two weeks</u> , how often have you been bothered by the following problems?	Not at all	Several days	More than half the days
1. Feeling nervous, anxious, or on edge	0	1	2
2. Not being able to stop or control worrying	0	1	2
3. Worrying too much about different things	0	1	2
4. Trouble relaxing	0	1	2
5. Being so restless that it is hard to sit still	0	1	2
6. Becoming easily annoyed or irritable	0	1	2
7. Feeling afraid, as if something awful might happen	0	1	2

Column totals _____ + _____ + _____

Total score

Beck's Anxiety Inventory

Numbness or tingling

Feeling hot

Worriedness in legs

Unable to relax

Fear of worst happening

Dizzy or lightheaded

Heart pounding / racing

Unsteady

Terrified or afraid

Nervous

Feeling of choking

Hands trembling

Shaky / unsteady

Fear of losing control

Difficulty in breathing

Fear of dying

Scared

Indigestion

Faint / lightheaded

Face flushed

Hot / cold sweats

	A	B	C	D	
1	Measure	Cohort	Age (Range)	Age (Year)	Low Mood
2	SF-36 (10 items)	ALSPAC	20s	18	6. Have you felt downhearted and person
3	MFQ	ALSPAC	20s	18	1. I felt miserable or unhappy / 3. felt happy
4	SF-36 (10 items)	ALSPAC	20s	21	6. Have you felt downhearted and person
5	MFQ	ALSPAC	20s	21	1. I felt miserable or unhappy / 3. felt happy
6	MFQ	ALSPAC	20s	22	1. I felt miserable or unhappy / 3. felt happy
7	Malaise Inventory (24-item version)	NCDS	20s	23	3. Do you often feel depressed?
8	MFQ	ALSPAC	20s	23	1. I felt miserable or unhappy / 3. felt happy

Browse or search studies and measures

Search for studies, measures, instruments...



Studies found: 61

Filters

Related measures >

Complementary data >

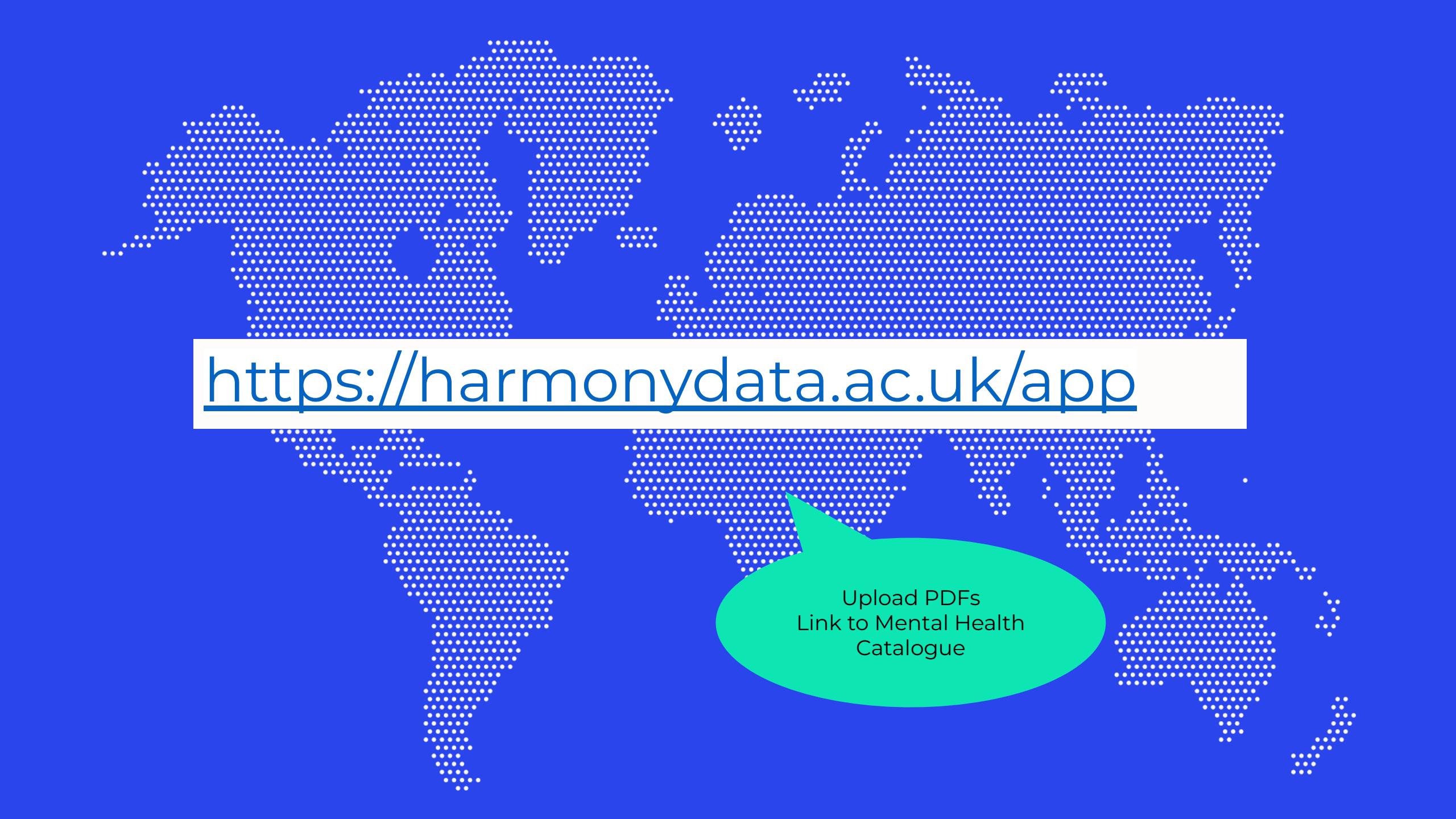
Study design >

Sample characteristics >

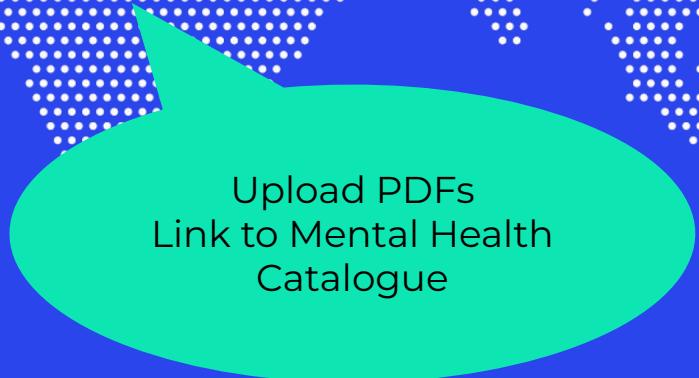


ADVANCE (Armed Services Trauma Rehabilitation Outcome Study)

The Armed Services Trauma Rehabilitation Outcome (ADVANCE) Study is a prospective longitudinal cohort study of the long-term physical and psychosocial outcomes of UK military personnel who sustained severe combat-related injuries during the Afghanistan conflict. ADVANCE aims to study the incidence of a wide variety of outcomes, including cardiovascular disease, musculoskeletal disease, mental health, functional and social outcomes, quality of life, employment, and more. Data are collected using various methods, including self-report questionnaires and clinical assessments.



<https://harmonydata.ac.uk/app>



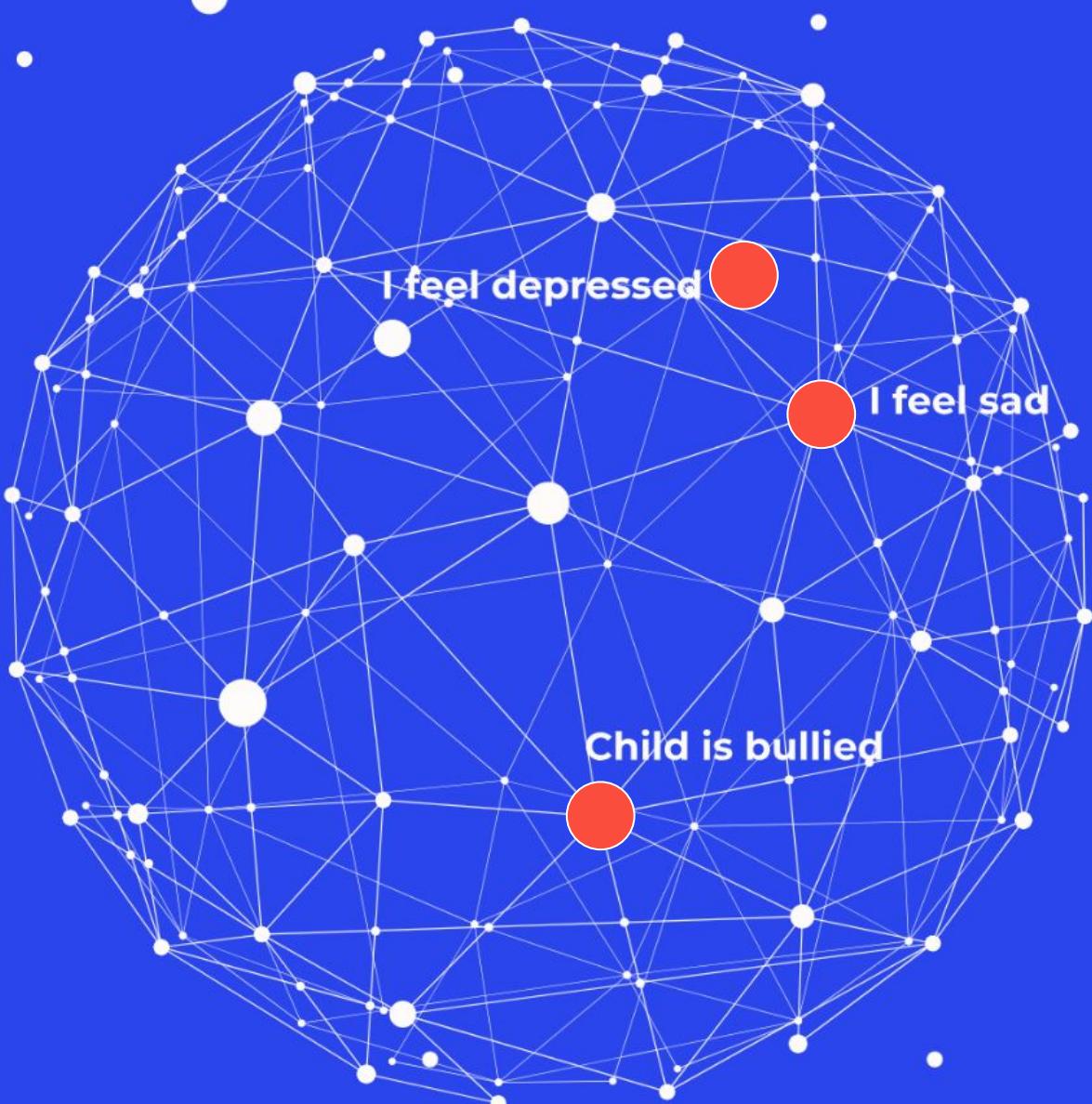
Upload PDFs
Link to Mental Health
Catalogue

How does Harmony work?

1. Process the PDF to get the question items out
2. Convert all questions to sentence embeddings
3. Calculate the cosine similarity



[Open in Colab](#)



Sentences have a related meaning

Vectors point in a similar direction

Cosine score is high – close to 1

Sentences have unrelated meaning

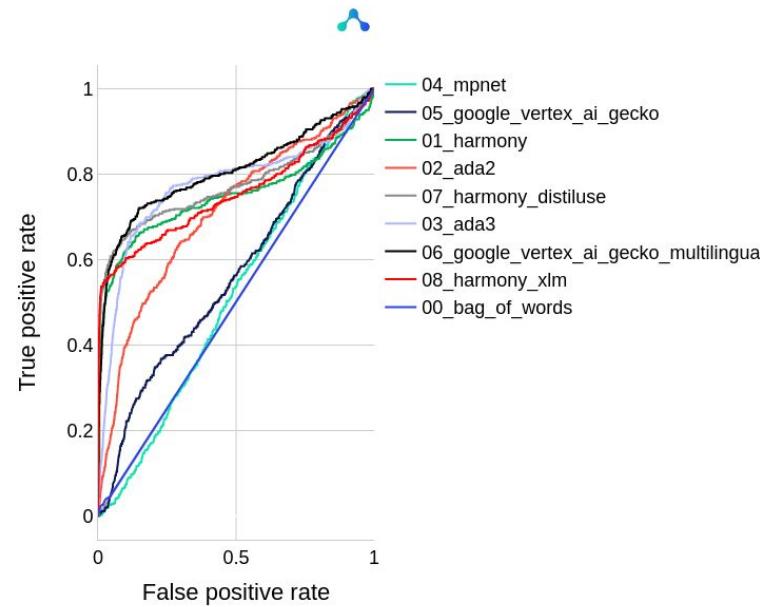
Vectors are orthogonal

Cosine score is low – closer to 0

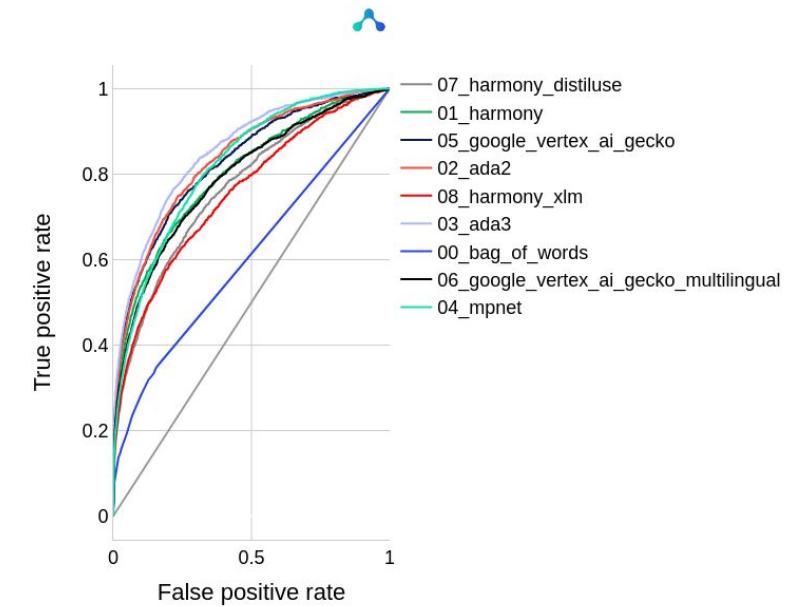


Evaluating Harmony

ROC on GAD 7 multilingual dataset



ROC on McElroy et al childhood dataset



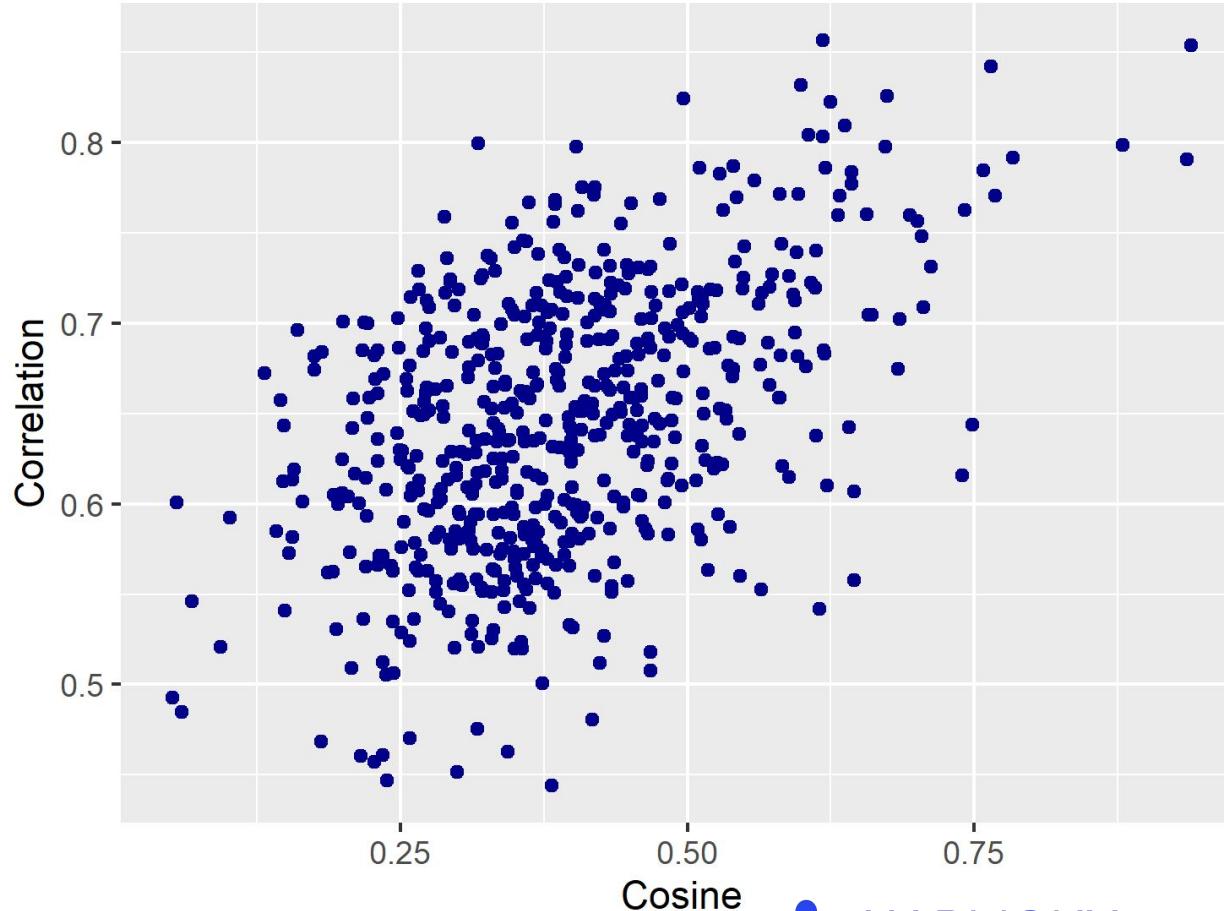


Real correlations?

	A	B	C					
1	Questionnaire	Item number	Content					
2	IDQ							
3	IDQ							
4	IDQ							
5	IDQ							
6	IDQ							
7	IDQ							
8	IDQ							
9	IDQ							
10	IDQ							
11	IAQ							
12	IAQ							
13	IAQ							
14	IAQ							
15	IAQ							
16	IAQ							
17	IAQ							
18	IAQ							
19	DHQ							
SECTION								
			A		B	C	D	E
			1 Supplementary File 2. Correlaiton and cosine coefficients for item pairs		from	to	spearman	cosine
					1	10	0.719538559	0.61149627
					2	11	0.719244021	0.445720732
					3	12	0.731182941	0.711875081
					4	13	0.665979411	0.571581244
					5	14	0.703580795	0.511955619
					6	15	0.709961188	0.297138691
					7	16	0.691486691	0.50184983
					8	17	0.608636306	0.259960353
					9	18	0.738485659	0.37026453
					10	19	0.798613012	0.879561961
					11	2	0.824578169	0.496798843
					12	20	0.573991369	0.572757006
					13	21	0.628075543	0.655307412
					14	22	0.625625193	0.354875147
					15	23	0.718334673	0.681931853
					16	24	0.683265023	0.334293664
					17	25	0.587003714	0.53709048



Real correlations?



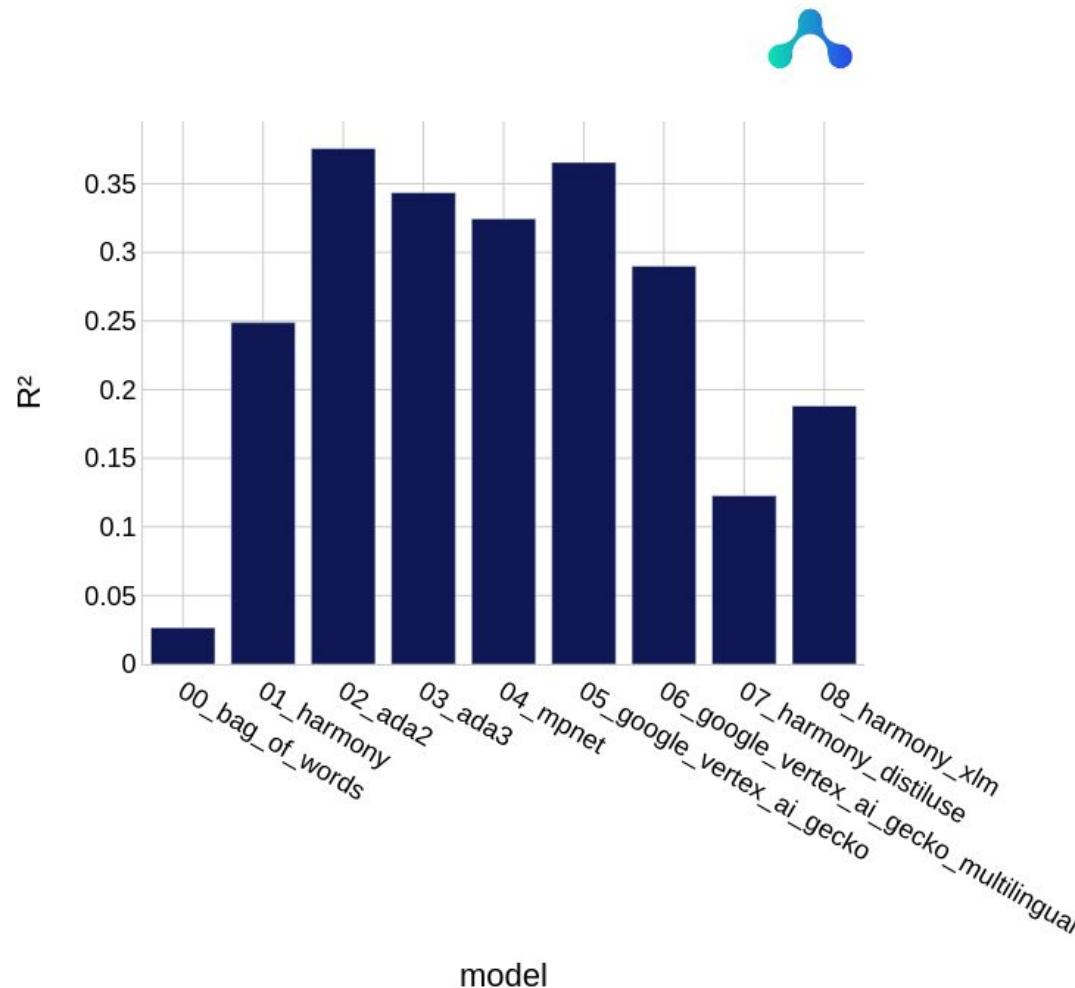
MCElroy et al,

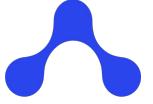
Using natural language processing to facilitate the harmonization of mental health questionnaires: a validation study using real-world data



Real correlations?

R^2 on McElroy et al cosine correlation dataset





Open source

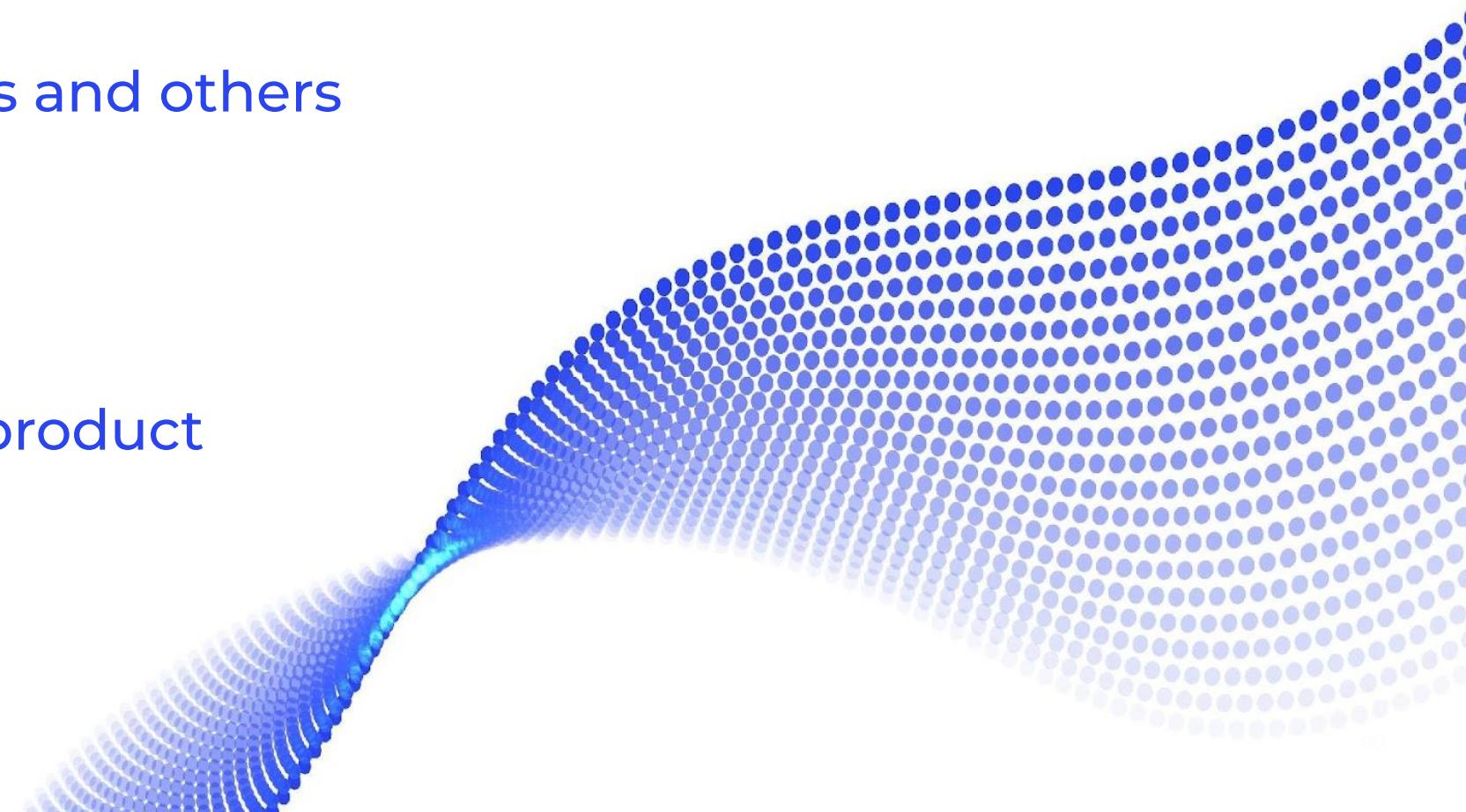
<https://github.com/harmonydata>

Free for psychologists and others
around the world

MIT License

It's not a monetised product

Hackathons



Discord

discord.gg/harmonydata





Uses outside psychology?

Age of HIV-1 uninfected person
Please complete this section if you have any physical health problems
All details will be treated in confidence
Do you have or suspect you have:
Rheumatic fever?
Any heart complaints?
Diabetes?
Epilepsy or fainting fits?
Chronic bronchitis?
Hepatitis?
Excessive bleeding?
High blood pressure?
Any other serious health problem?
Do you smoke?

Find your local integrated care board (ICB)

Integrated care boards (ICBs) replaced clinical commissioning groups (CCGs) in the NHS in England from 1 July 2022.

East of England

- [NHS Bedfordshire, Luton and Milton Keynes ICB](#)
- [NHS Cambridgeshire and Peterborough ICB](#)
- [NHS Hertfordshire and West Essex ICB](#)
- [NHS Mid and South Essex ICB](#)
- [NHS Norfolk and Waveney ICB](#)
- [NHS Suffolk and North East Essex ICB](#)

London





Uses outside psychology?

- Market research? (Surveys about new products?)
- Pharma? (Informed Consent Forms?)
- Finance?
- Legal?
- National Archives?
- Your industry?





Next steps

We are working on linking to external data sources such as

- Mental Health Catalogue
- UK Longitudinal Linkage Collaboration

Improve PDF parsing

Improve question item matching

Link to more databases, browser plugins, more collaborations...





Open source tasks

We are always looking for
collaborators.
Ideas include

- Create browser plugin
- Integration with more platforms
- PDF report export function
- Ability to harmonise data (can we create a Notebook on the fly?)





Hamilton Anxiety Scale

Rates level of anxiety based on clinical questions.

Favorite Share tool

When to Use ▾

Why Use ▾

About the Creator



Dr. Max Hamilton

Ctrl+C

1. Anxious mood: Worries, anticipation of the worst, fearful anticipation, irritability, restlessness, tension, panic attacks, avoidance, hypervigilance, difficulty concentrating, muscle tension, sweating, tachycardia, hyperventilation, dry mouth, nausea, diarrhea, constipation, urinary frequency, sexual dysfunction, fatigue, and sleep disturbances.

Copy

Copy link to highlight

Search Google for "1. Anxious mood: Worries, anticipation of the worst, fearful anticipation, irritability, restlessness, tension, panic attacks, avoidance, hypervigilance, difficulty concentrating, muscle tension, sweating, tachycardia, hyperventilation, dry mouth, nausea, diarrhea, constipation, urinary frequency, sexual dysfunction, fatigue, and sleep disturbances."

Print...

Translate selection to English

Open in reading mode



Send to Harmony (Ctrl+Shift+H)

Inspect

Moderate

Severe

Very Severe

2. Tension: Feelings of tension, fear, apprehension, nervousness, restlessness, difficulty relaxing, muscle tension, sweating, tachycardia, hyperventilation, dry mouth, nausea, diarrhea, constipation, urinary frequency, sexual dysfunction, fatigue, and sleep disturbances.



**Other projects
with difficult
PDF parsing...**





Clinical trial risk tool



Trial is for condition [explain](#)

HIV

Trial phase [explain](#)

2

Has the Statistical Analysis Plan been completed?

[explain](#)

Yes

Has the Effect Estimate been disclosed? [explain](#)

Yes

Number of subjects ● ⚠ Low confidence! [explain](#)

90

Possible sample sizes found: 90, 45, 41

Sample size tertile: 0 (small trial) [set values of tertiles](#)

Number of arms [explain](#)

2

Countries of investigation ● [explain](#)

×  United States

Clinical trial risk tool

<https://app.clinicaltrialrisk.org>

Tool extracts key features from trial text and puts them into risk model

Risk of uninformativeness

MEDIUM



Protocol: 26_NCT02263326_Prot_SAP_ICF_000.pdf
(47 pages)

[Export report as PDF](#)

[View log of the analysis](#)



Thank you!

Question, Suggestions and (maybe) answers

@harmony_data
@fastdatascienc1

<https://harmonydata.ac.uk/app/>

<https://discord.gg/harmonydata>
<https://github.com/harmonydata>



Fast Data Science



CENTRE FOR
LONGITUDINAL
STUDIES

