

LESSONS LEARNED FROM A SECONDARY ANALYSIS USING NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING FROM A LIFESTYLE INTERVENTION

BACKGROUND:
Fidelity automation and prediction of lifestyle behavioral outcomes through *natural language processing (NLP)* and *machine learning (ML)* were met with challenges related to the *re-usability of the LIVES Study (GOG 0225) data*.

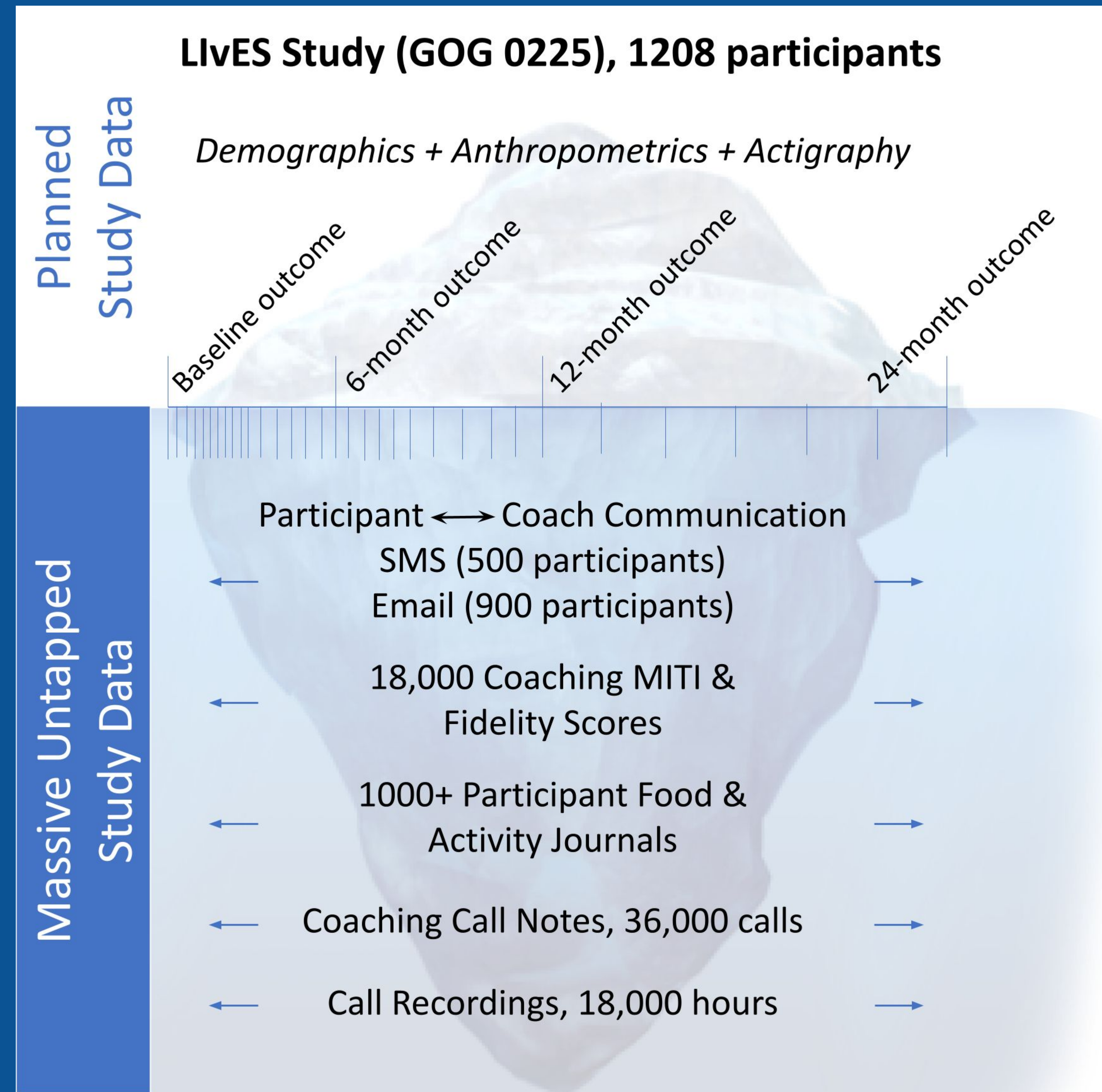
CHALLENGES: Numerous costly measures taken to re-use our data:

Challenge	Cost
Audio files and fidelity data not linked to outcomes	~240 hrs of manual work + data loss
Lack of planning for data derivatives	~170 hrs incl. re-annotation
Incomplete socio-cultural background data	Missed opportunities
Personally identifiable information	Missed opportunities
Consent	40 hrs of work + data loss

- LESSONS LEARNED:**
- Ensure unique identifiers link all related data
 - Plan for machine-friendly data structure and format
 - Co-collect data and derivatives
 - Collect culturally relevant data
 - Create data de-identification protocols for all data types
 - Consent all the people involved, not only participants

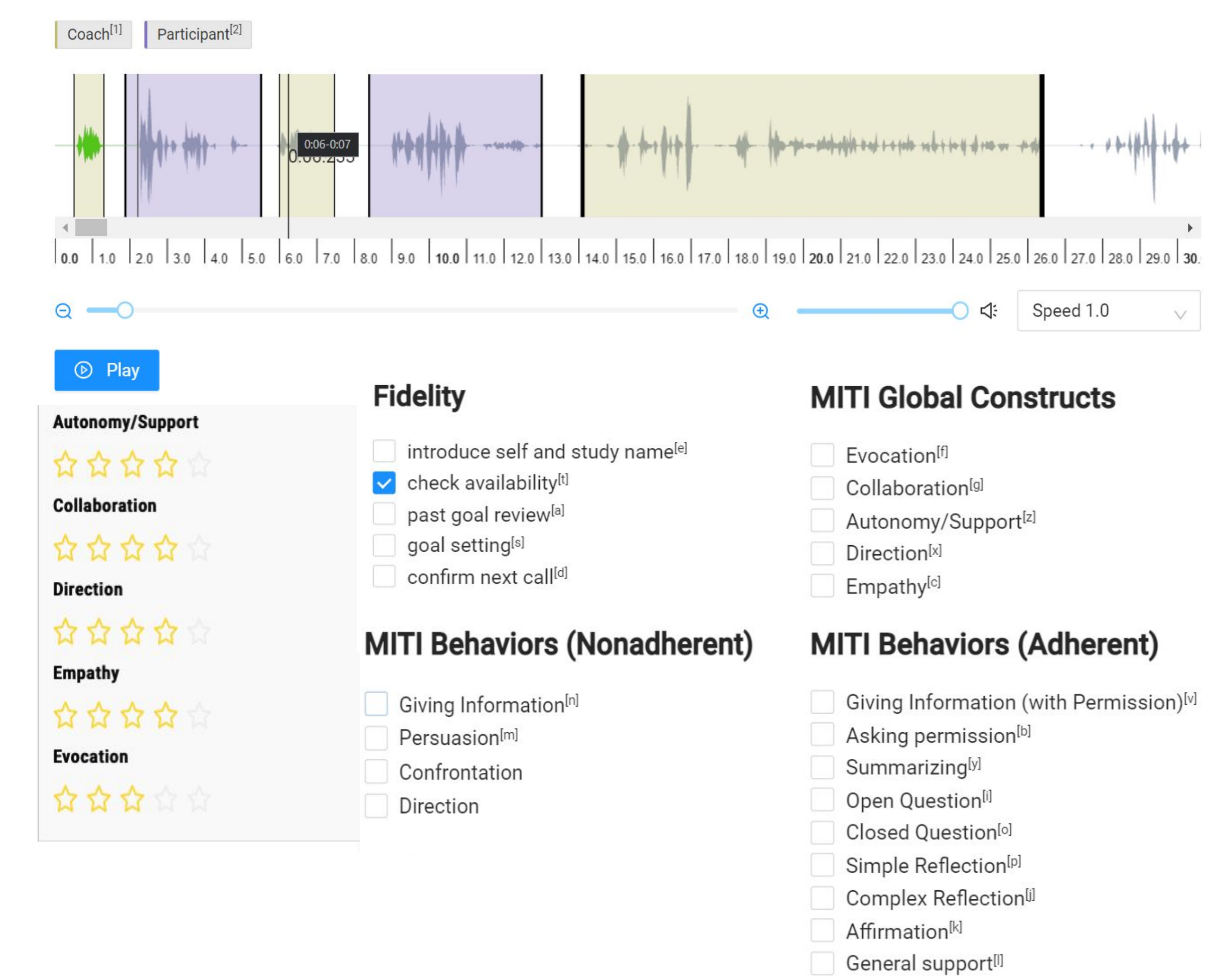


Comprehensive *data re-usability* is critical in the current ecosystem of machine learning and artificial intelligence



Actionable data management foresight can help extract value from non-traditional sources of data from behavioral health interventions

LIVES Study (GOG 0225) NCT00719303
BMISr UACC P30 CA023074
NLP/ML NIH/NCI 1R21CA256680-01



Natural Language Processing & Machine Learning compatible annotation interface

Draft Manuscript: Do Not Quote Without Author's Permission 27

Motivational Interviewing Treatment Integrity Code (MITI)
Coding Sheet Revised June, 2007
Tape # 007-0225-016 Coders: [Signature] Date: 2/15/18

Global Ratings

Evocation	How do you want to tackle this?	1 Low 2 3 4 5 High
Collaboration	How would this help you lose weight?	1 Low 2 3 4 5 High
Autonomy/Support		1 Low 2 3 4 5 High
Direction	missed opportunity, but covered on topic	1 Low 2 3 4 5 High
Empathy	Good listening	1 Low 2 3 4 5 High


Behavior Counts

Giving Information		1	
MI Adherent	Asking permission, affirm, emphasize control, support.	1111	1
MI Non-adherent	Advise, confront, direct.		
Question (subclassify)	Closed Question	111	
	Open Question	111	
Reflect (subclassify)	Simple	111	
	Complex		
	TOTAL REFLECTIONS:	3	

First sentence: Is this Ann? This is Sophie
Last sentence: Mrs husband had just had surgery for

Natural Language Processing & Machine Learning incompatible annotation interface

CONCLUSION:
To take advantage of machine learning, natural language processing and artificial intelligence, behavioral interventions should engage the support of a data scientist in the study design planning stage and throughout data collection.

 Sarah Freylersythe,^{1,4}
Rebecca Sharp,⁷ John
Culnan,⁵ Damian Y. Romero
Diaz,³ Yiyun Zhao,⁵ G.
Hagan Franks,⁶ Remo
Nitschke,⁵ Steven J.
Bethard,² Tracy E. Crane⁴

