ARTIFICIAL INTELLIGENCE AND PSYCHIATRY: A FOCUS ON DIAGNOSTIC AUTOMATION

Some preliminary work experiences

Gaël DIAS @ Sorbonne

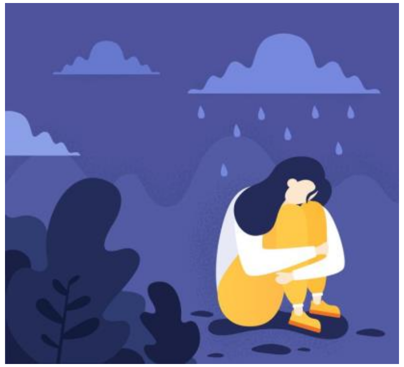
joint work with
Navneet AGARWAL, Mohammed HASANUZZAMAN, Arbaaz QURESHI, Kirill MILINTSEVICH, Valentin RENIER, Soumaya SABRY, Sriparna SAHA, Kairit SIRTS, and more to come ;)

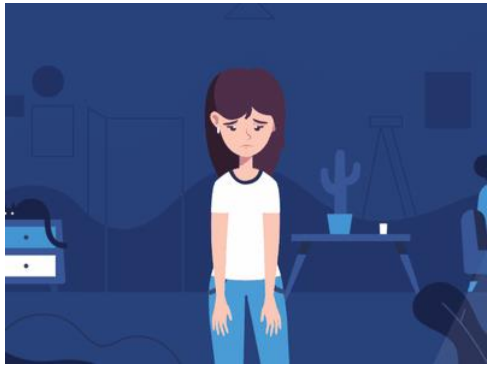
University of Caen Normandie - CNRS - GREYC UMR 6072 Research Laboratory
gael.dias@unicaen.fr

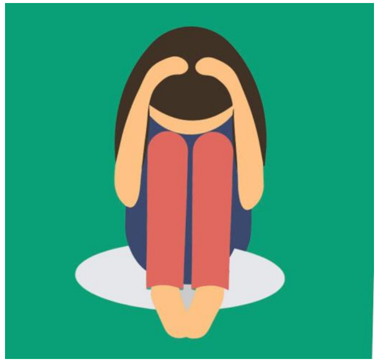
GREYC Electronics and Computer Science Laboratory

Normandie Université UNICAEN ENSI CAEN ÉCOLE PUBLIQUE D'INGÉNIEURS CENTRE DE RECHERCHE cnrs

Outline







Mental Health

- The world is experiencing a mental health crisis.
- It is estimated that 970 million people worldwide had a mental or substance use disorder in 2017, of which 284 million showed anxiety disorders and 264 million suffered from depression, mostly affecting females (source Forbes).
- What's next after/during the COVID-19? Unemployment, divorce, etc.
- The critical shortfall of psychiatrists and other mental health specialists to provide treatment exacerbates this crisis. In the Ain region in France, the supply of psychiatric care is half the national average, i.e. 9 psychiatrists for 100,000 inhabitants (source Le Progrès). This shortage of doctors results in less frequent appointments and practitioners who no longer take new patients.
- This crisis is even more exacerbated in France, where Psychiatry has been defined as the "parent pauvre de la médecine" (i.e. the poor relative of medicine) by the French Ministry Agnès Buzyn in 2018. In particular, she stated that "*Psychiatry is a discipline of the future, but the organization of mental health care and its place in the society are not up to the task [...]. Prevention is insufficient, and diagnosis too late [...]. I make it a health priority*". (Source Science et Avenir).

Mental Disorders

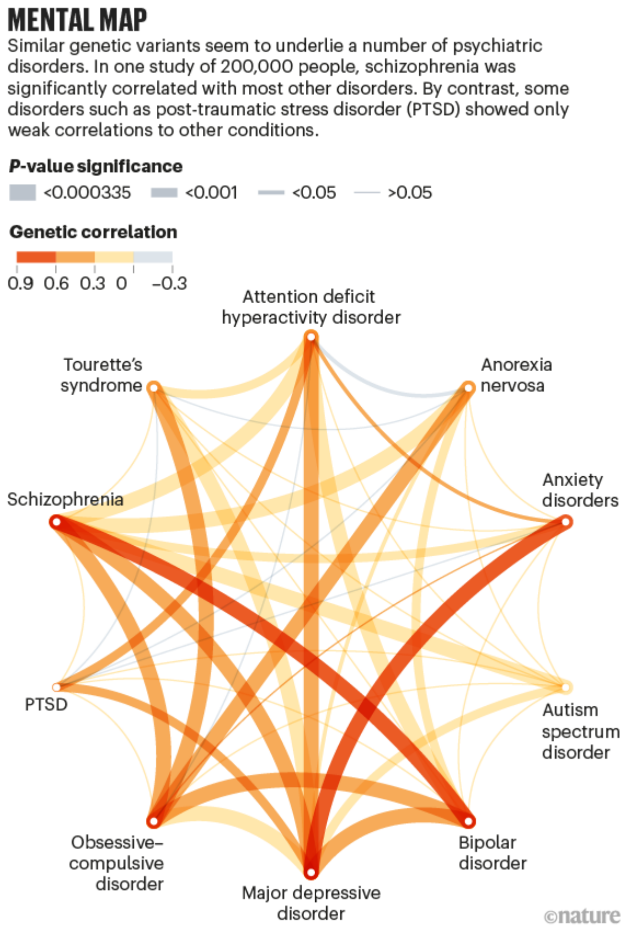
- Anxiety disorders,
- Bipolar and related disorders,
- Depressive disorders,
- Disruptive, Impulse-control, and Conduct disorders,
- Dissociative disorders,
- Feeding and eating disorders,
- Gender dysphoria,
- Obsessive-compulsive and related disorders,
- Personality disorders,
- Trauma and stressor-related disorders,
- Schizophrenia spectrum and other psychotic disorders,
- etc.

Many Different Symptoms

- Apathy,
- Avoidance,
- Excessive fear or uneasiness,
- Feeling of disconnection,
- Increased sensitivity,
- Mood changes,
- Problems thinking,
- Significant tiredness,
- Sleep or appetite changes,
- Withdrawal,
- etc.

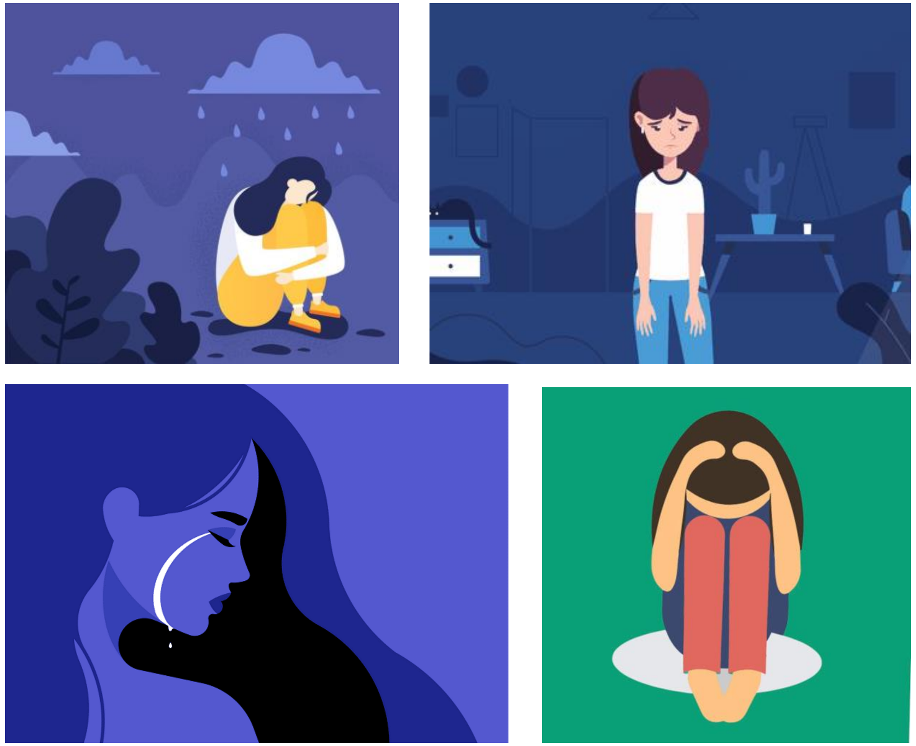
Depression

- Depression is characterized by chronic low mood, low self-esteem and loss of interest.
- Depression is a disabling condition that can impact family, school or work. In the most severe cases, depression is characterized by a high suicide rate.
- The causes of depression are multiple and not well understood: e.g. genetic predisposition, traumatic experiences, inability to cope with rejection or failure.
- The diagnosis of depression is based on the patient's personal feelings, the behavior perceived by those around and the results of psychological examination.
- The diagnosis of depression is complex due to:
  - the high rate of comorbidity,
  - the subjectivity of the examinations,
  - the non-regular therapeutic follow-up,
  - the patient coverage of symptoms.



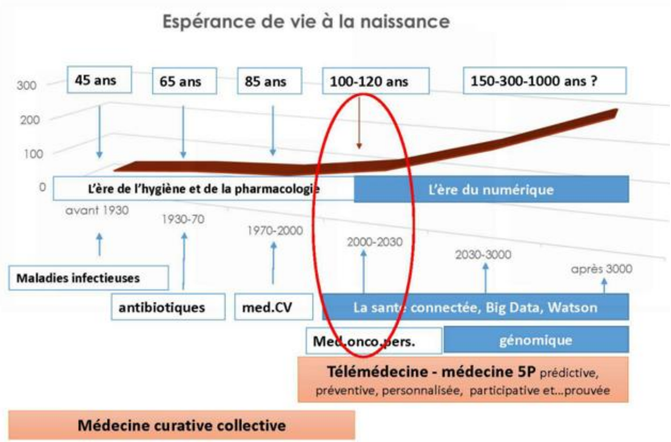
Outline

1. *Mental Health and Depression*
2. **6P Medicine**
3. **Interesting Iniatives**
4. **Computer-aided Diagnosis**
   i. *Multimodality*
   ii. *Emotionality*
   iii. *Gender-awareness*
   iv. *Dialogue structure*
   v. *Symptom-based diagnosis*
5. **A Favourable Research Environment**

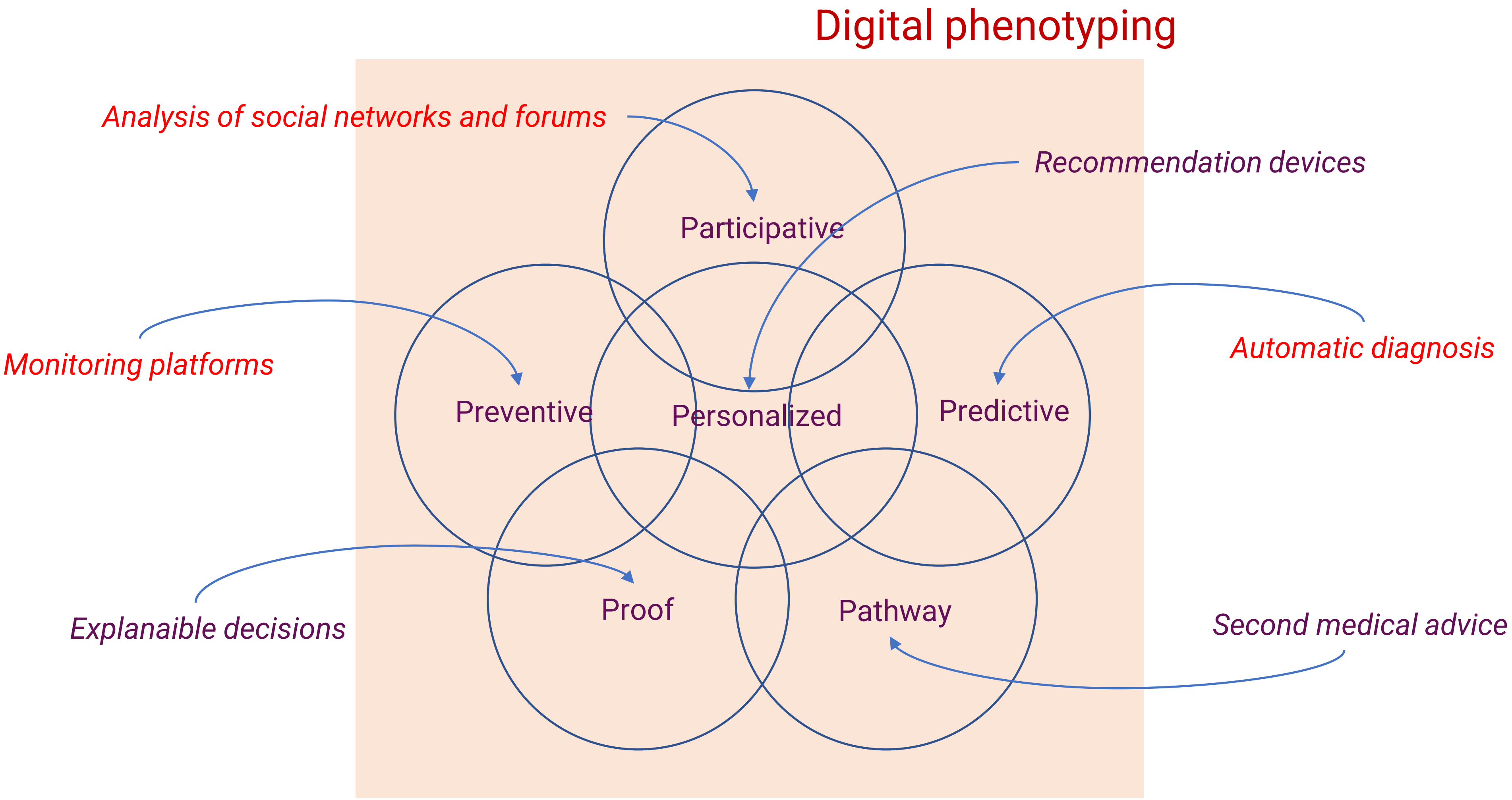


6P Medicine

- 1P - Personalized: Personalized medicine consists of adapting a medical treatment according to the individual characteristics of a patient.
- 2P - Preventive: Preventive medicine focuses on wellness, and consists of measures taken for disease prevention.
- 3P - Predictive: Predictive medicine is a branch of medicine that aims to identify patients at risk of developing a disease.
- 4P - Participative: Medicine should be participatory, leading patients to be more responsible for their health and care.
- 5P - Proof: Medicine must be based on evidence of medical service to patients, especially when it relies on connected health and telemedicine.
- 6P - Pathway: Coordinating multiple interventions (medical, social, occupational medicine, etc.) such that the healthcare pathway is progressively articulated, according to the pathology and its evolution.

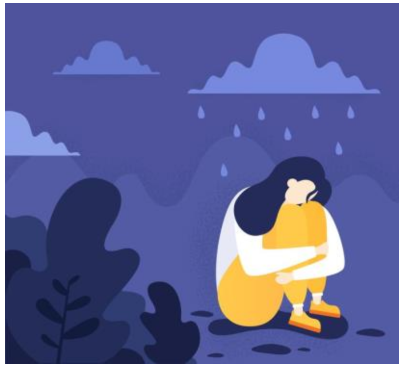
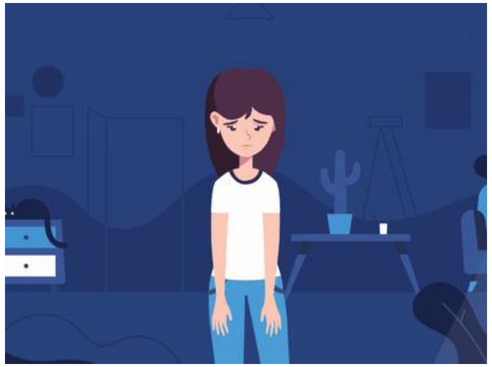


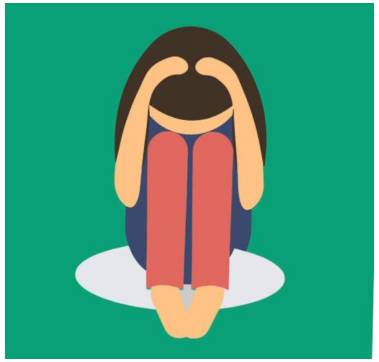
6P Medicine and AI



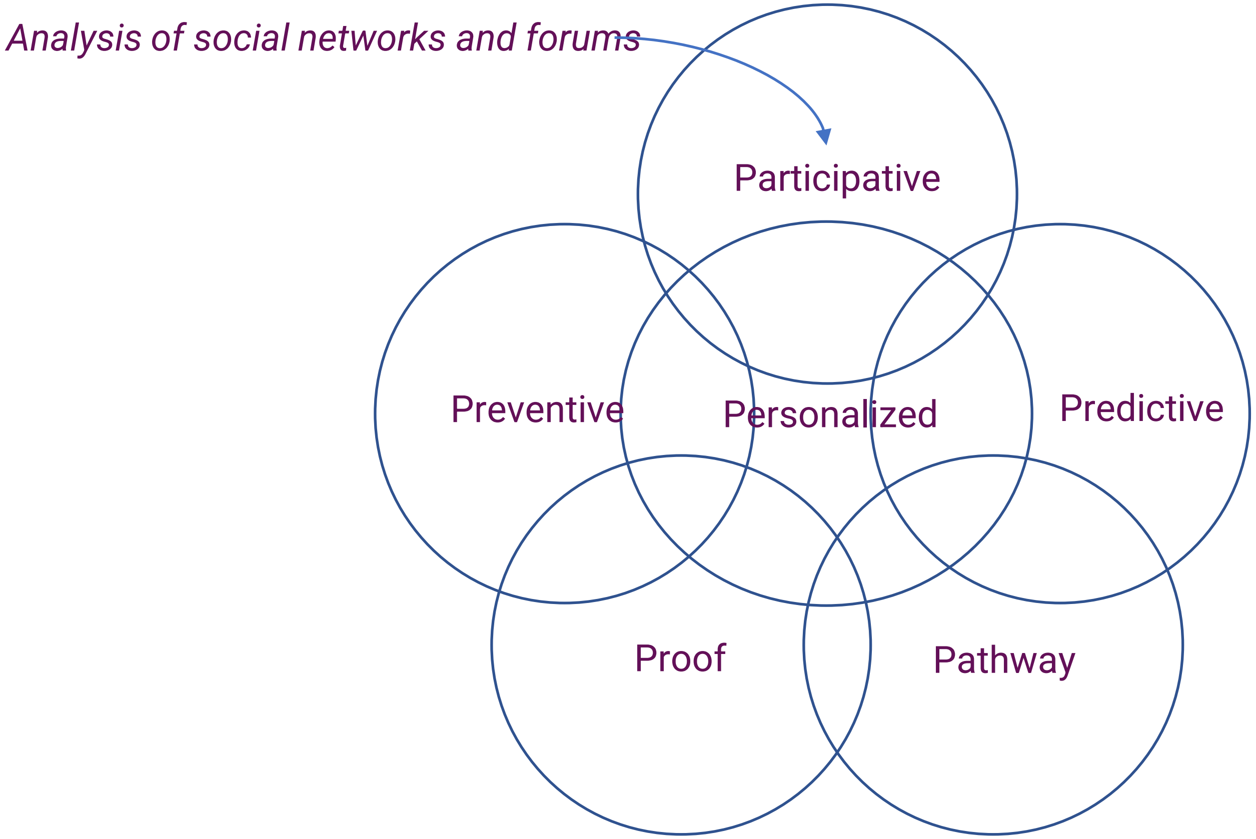
Outline

1. *Mental Health and Depression*
2. *6P Medicine*
3. **Interesting Iniatives**
4. **Computer-aided Diagnosis**
   1. *Multimodality*
   2. *Emotionality*
   3. *Gender-awareness*
   4. *Dialogue structure*
   5. *Symptom-based diagnosis*
5. **A Favourable Research Environment**

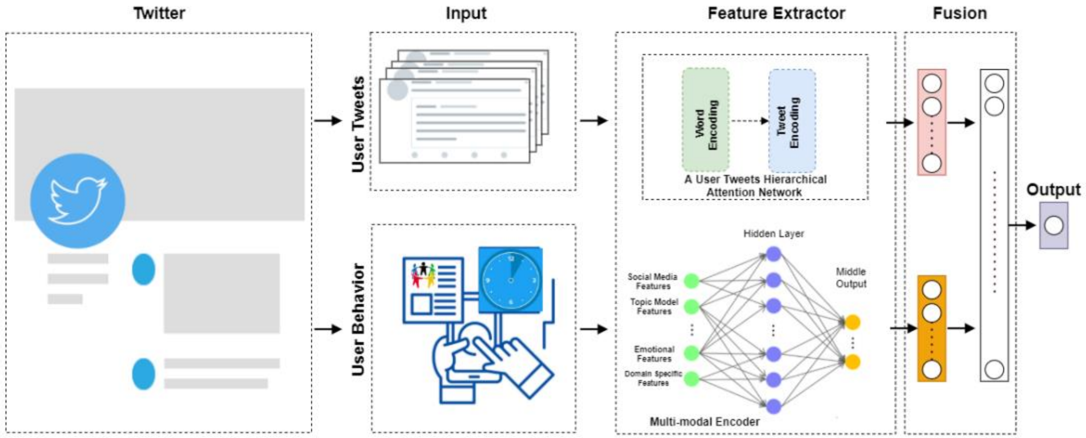


Computer-aided Diagnosis



Social Network Analysis

- Social networks are an important support for Participative medicine, which automatic analysis might allow Preventive/Predictive actions.
- It is common for people who suffer from mental health problems to often disclose their feelings and their daily struggles with mental health issues on social media as a way of relief.
- Twitter, Reddit, Doctissimo, to name but a few platforms have become an excellent resource to automatically discover people who are under depression.
- [Zogan et al., 2021] propose a depression detection framework by tackling textual, behavioral, temporal, and semantic modalities.



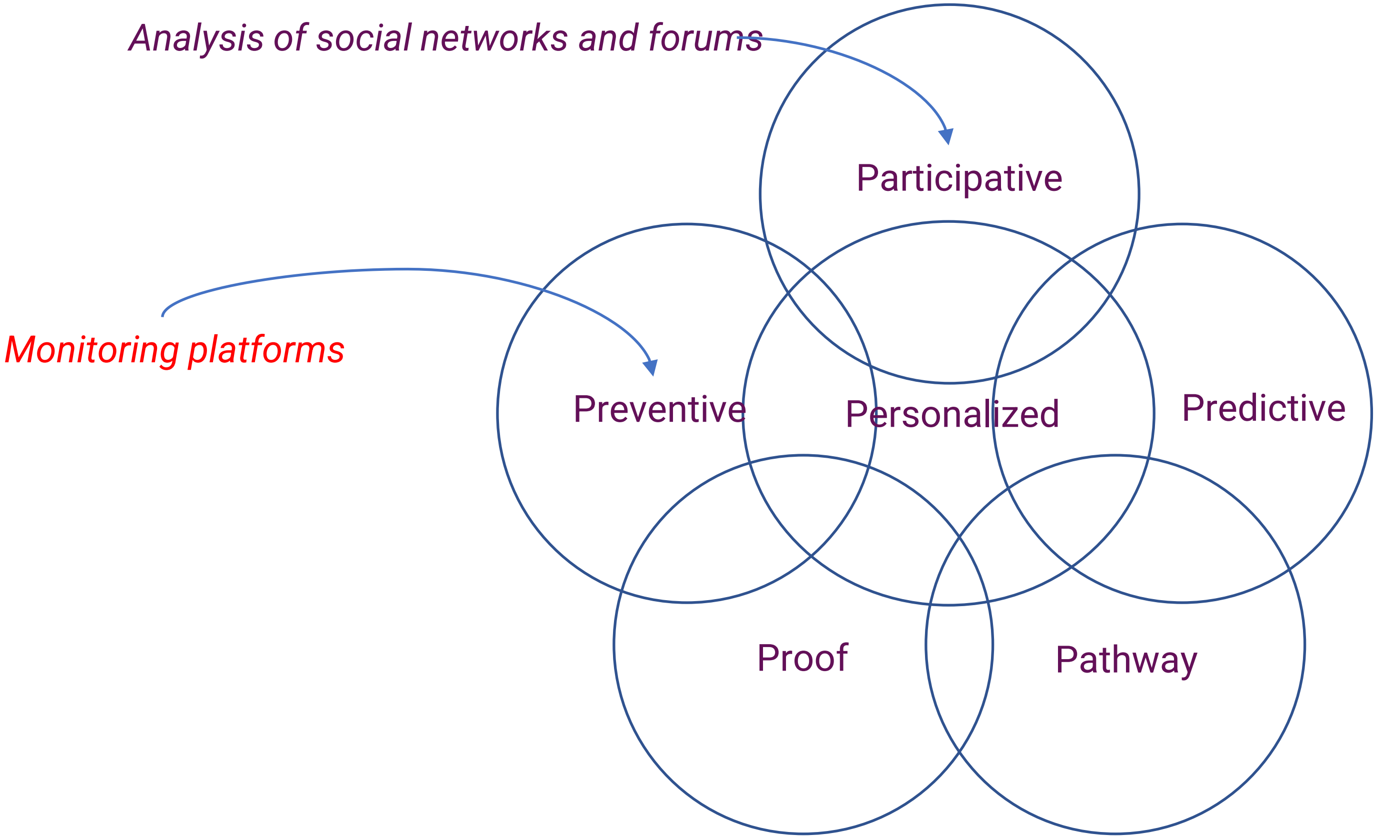
Social Network Analysis

- [Losada & Gamallo, 2018] propose to analyze and improve current language resources for identifying signs of depression on Reddit.
- Other lexicons: Pedesis (2012) obtained from the web, Choudhury (2013) based on Twitter analysis, Schwartz (2014) focused on Facebook posts, etc.
- They propose to expand existing lexicons with selected terms following distributional and paradigmatic-based models, and thesaurus-based models.
- Their Rocchio based experiments show that the resulting lexica are effective at identifying signs of depression in a non-supervised way.

accelerate adsorb affect alleviate anger ask avoid beat bestow blotched bruise cancel capture carry cause cdot characterise characterize clinch collapse colour confront conquer convert convince cry decline defeat define delay denote depopulate derive destroy detect devastate devote diminish disappear disappoint divide elongate emit encircle enclose encourage enlarge erode evaporate evoke evolve exacerbate exclude exercise extract facilitate fade fill finish flank flatten fleck focus foil forward grab grieve halt hamper hawthorn heal hinder hope impede imply impress induce infuse inject innervate invade ionize isolate kill leach metabolize minimize opt orange-red outflank outrage overhang owe oxidise oxidize pacify peasantry penetrate pertain plan postpone pray prepare present prevent protrude ravage react refer relate remove repel repulse reschedule respond revere reward satisfy schedule seedling seep send separate sharpen shock shower slate soothe speckle stop streak strive subdue subjugate submit surprise surround swell taper tell thwart ting transform traverse treat tremble turn urinate vaporize venerate vine vomit wait wane wield win wish worship yearn

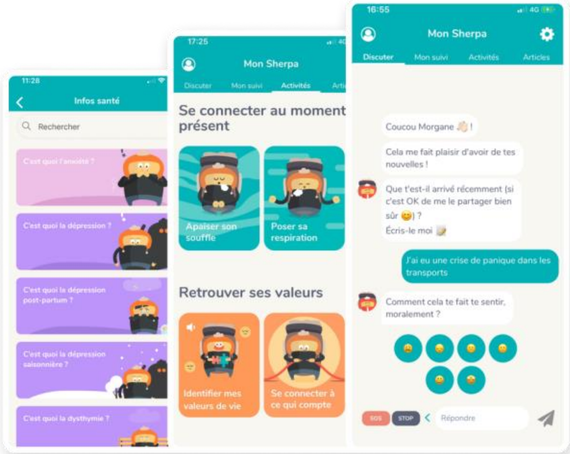
Table 7 New words included in the Pedesis lexicon by the DE expansion method

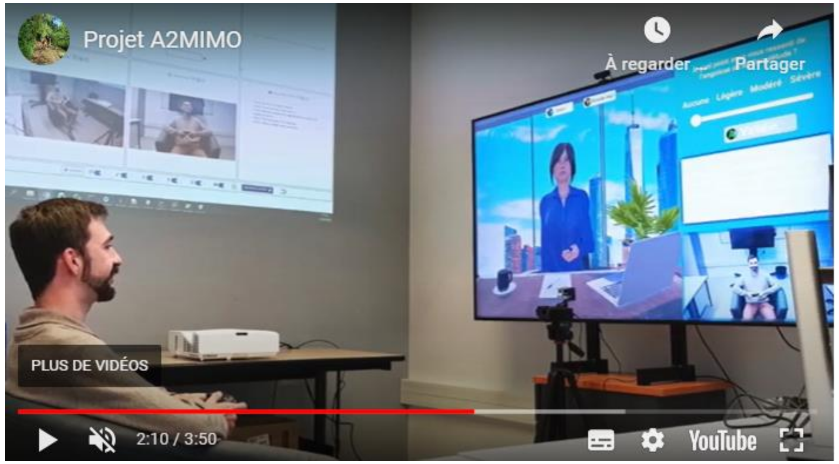
Computer-aided Diagnosis



Monitoring Platforms and (Embodied) Chatbots

- A chatbot is a system that is able to converse and interact with human users using spoken, written, and visual languages (embodied).
- Chatbots can be useful preventive tools for individuals who are reluctant to seek mental health advice due to stigmatization.
- [Abd-alrazaq et al., 2019] studied 41 different embodied and non-embodied chatbots. Most tackle depression and autism.
- Among other scientific issues, therapeutic alliance is the key factor for the success of chatbots and ECAs.

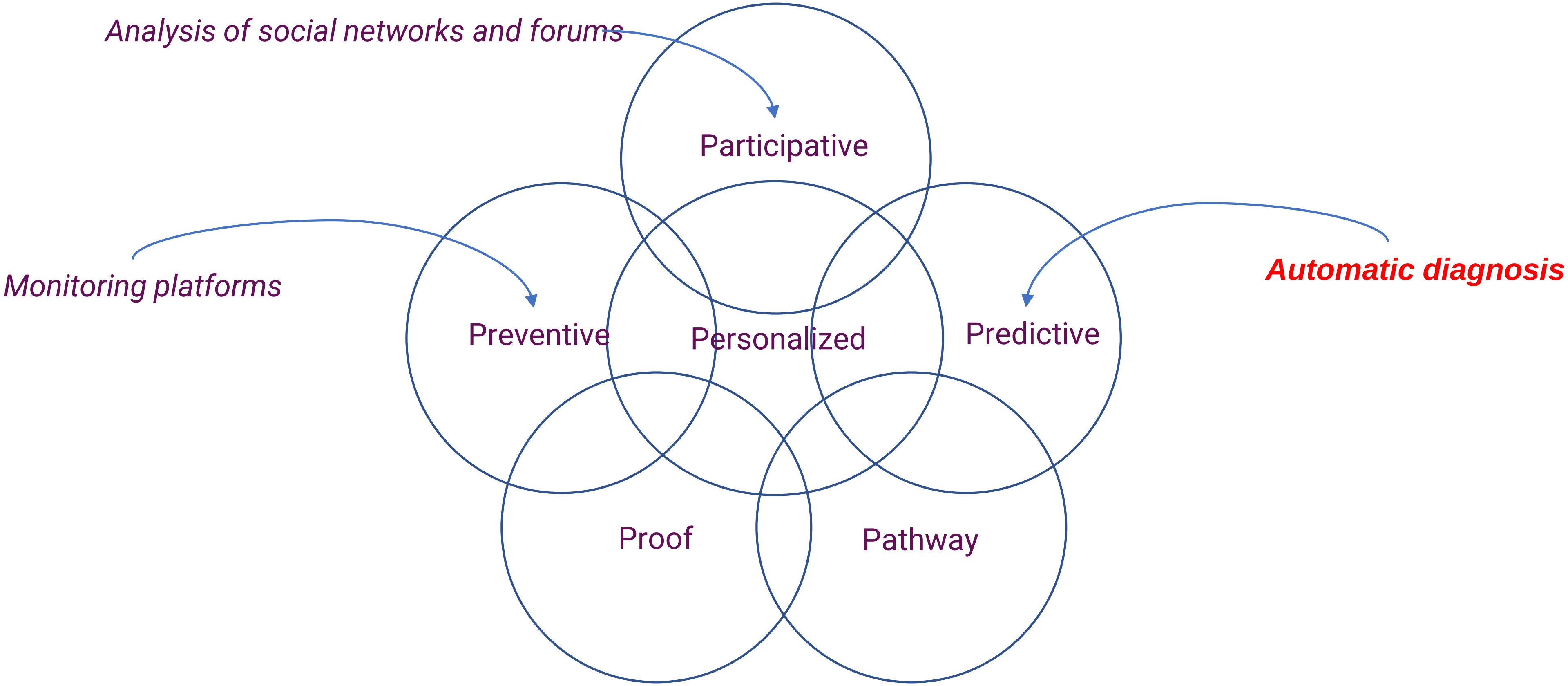




Data Sets and Related Events

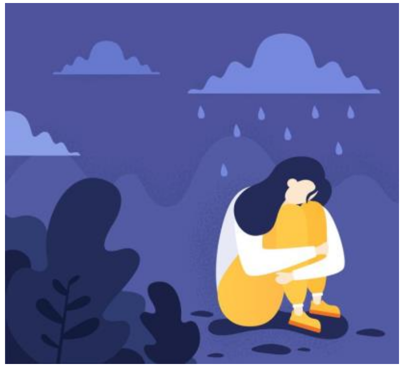
- The major issue with mental health applications is the availability of datasets. Most datasets are not available for reproducibility.
- Some very few exceptions for clinical interviews:
  - DAIC-WOZ [Gratch et al. 2014].
  - General Psychotherapy Corpus [Alexander Street Press?].
  - Audio-visual Depressive Language Corpus [AVEC 2013].
  - Bipolar Disorder Corpus [AVEC 2018] – Turkish language / Bipolarity.
- More exist which are based on social networks :
  - Research on Depression in Social Media [Rissola et al. 2020].
  - Early Detection of Depression [eRisk 2017].
  - CLPsych dataset [Milne et al., 2016] – Risk, Red, Amber, Green / Depression and PTSD.
  - Early Detection of Signs of Anorexia [eRisk 2018].
  - Suicide Watch [Shing et al. 2018].
  - And certainly many others …

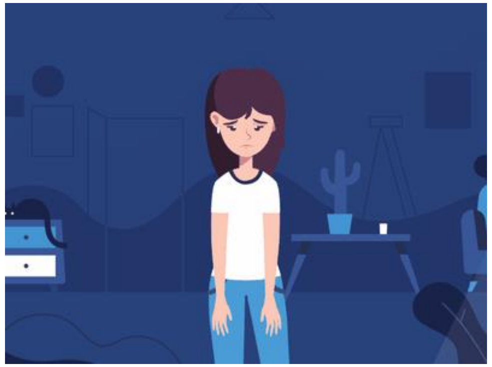
Computer-aided Diagnosis

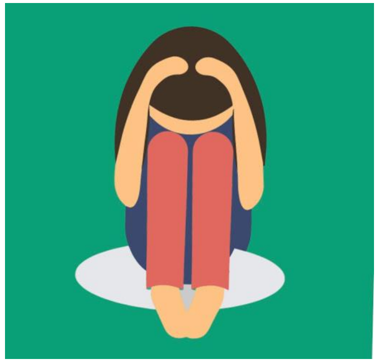


Outline

1. *Mental Health and Depression*
2. *6P Medicine*
3. *Interesting Iniatives*
4. **Computer-aided Diagnosis**
   i. *Multimodality*
   ii. *Emotionality*
   iii. *Gender-awareness*
   iv. *Dialogue structure*
   v. *Symptom-based diagnosis*
5. **A Favourable Research Environment**







DAIC-WOZ Dataset

- The DAIC-WOZ dataset [DeVault et al. 2014] includes Wizard-of-Oz interviews, conducted by an animated virtual interviewer called Ellie, controlled by a human interviewer in another room.
- This share includes 189 sessions of interactions ranging between 7-33min (average of 16min).
- Each session is combined with a PHQ-8 questionnaire.

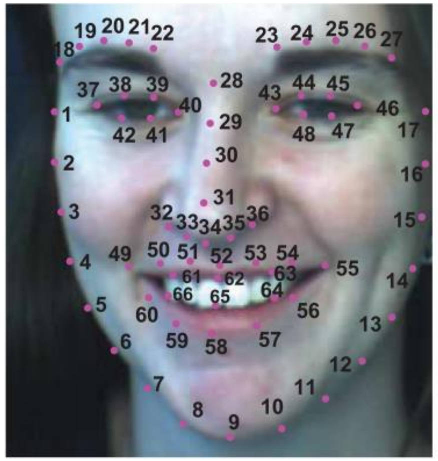
	Opening Rapport Building Phase
Ellie	*What are some things you really like about LA?* (top level question)
User	*I love the weather, I love the palm trees, I love the beaches, there's a lot to do here.*

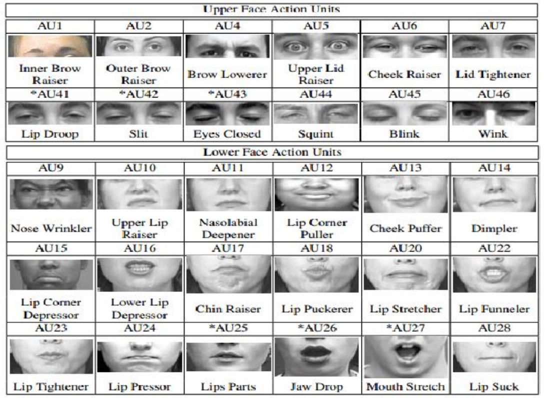
	Diagnostic Phase
Ellie	*Have you noticed any changes in your behavior or thoughts lately?* (top level question)
User	*Yes.*
Ellie	*Can you tell me about that?* (continuation prompt)
User	*I'm having a lot more nightmares now uh can't sleep have haven't really been eating uh trying to eat… I have to force down food um just feeling like an emotional wreck.*
Ellie	*I'm sorry to hear that.* (empathy response)
Ellie	*What are you like when you don't sleep well?* (follow-up question)
User	*Irritable, emotional, it just adds to my overall stress um [long pause]*
Ellie	*What…* (Ellie speaks after the participant's long pause)
User	*Can't concentrate uh I uh…* (the participant starts speaking while Ellie is speaking)
Ellie	*I'm sorry please continue.* (Ellie realizes that she has interrupted the participant and apologizes)

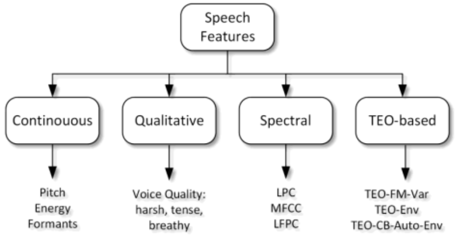
Questionnaire	0-1 day Not at all	2-6 days Several days	7-11 days More than half days	12-14 days Nearly everyday
1. Limited interest in doing work	15%	20%	40%	25%
2. Subjects with feeling of depression, hopelessness	20%	35%	15%	30%
3. Difficulty in sleeping or long sleep	40%	13%	25%	22%
4. Tiredness	20%	17%	35%	28%
5. Anorexia or excessive eating	14%	15%	37%	34%
6. Self bad feeling	10%	10%	30%	50%
7. Difficulty in concentration in work	30%	15%	15%	40%
8. Speaking or moving so slowly	23%	20%	25%	32%

Multimodal Estimation of PHQ-8

- In a patient-therapist interview, different signals shoud be combined for a correct diagnosis.
- Within the DAIC-WOZ dataset, the following signals are available:
  - Visual signals : expression of sadness, gaze escape, etc.
    - Facial Landmarks (FL), Head Pose (HP), Eye Gaze (EG), Action Unit (AU).
  - Speech signals : veiled voice, monotonous tone, etc.
    - Formant (FMT), COVAREP (COV).
  - Language signals : negative vocabulary, lack of perspective, etc.
    - Universal Sentence Encoder (TR).

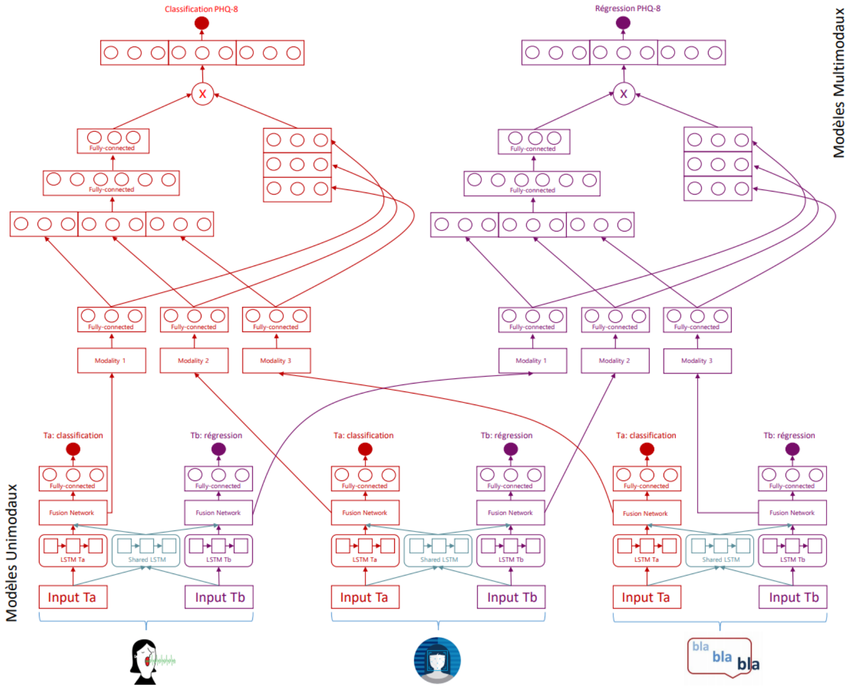


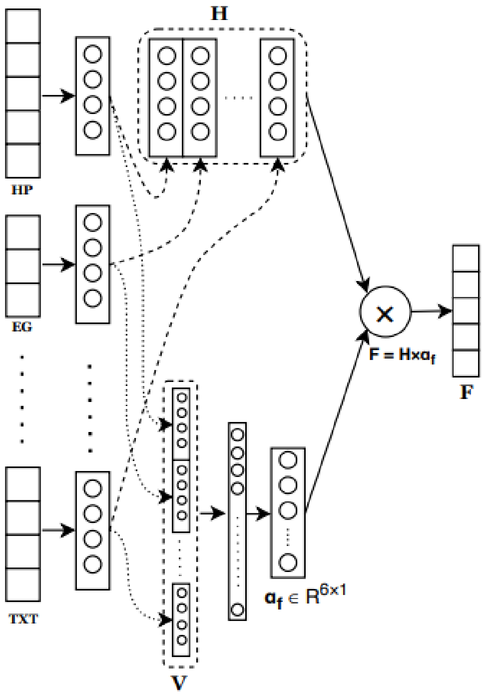


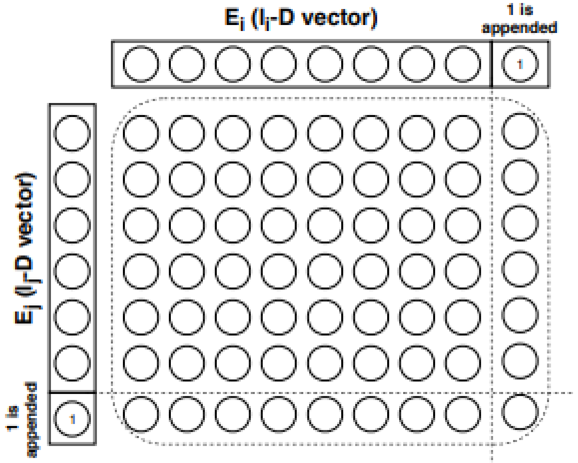


Multimodal Estimation of PHQ-8

- Combining classification and regression of depression estimators.
- An attention fusion network is used to combine inputs.
- Intra-modality inputs signals are combined with tensors.







An element of this matrix is the product of the numbers present in the corresponding row and column

Multimodal Estimation of PHQ-8

Table 1. Overall results. ST: Single Task, MT: Multitask, FS: Fully Shared, SP: Shared Private, DLC: Depression Level Classification, DLR: Depression Level Regression, HP: Head Pose, EG: Eye Gaze, AU: Action Units, COV: COVAREP, FMT: Formant, TXT: Text.

	Architectures	RMSE	MAE	Acc (%)	F-score
Unimodal	ST-DLR-HP	6.89	5.67	-	-
	ST-DLC-HP	-	-	54.54	0.41
	FS-MT-HP	6.75	**5.48**	**60.60**	**0.43**
	SP-MT-HP	**6.65**	5.53	54.54	0.42
	ST-DLR-EG	6.67	4.72	-	-
	ST-DLC-EG	-	-	54.54	0.37
	FS-MT-EG	**6.50**	**4.60**	**57.57**	**0.41**
	SP-MT-EG	6.59	5.16	57.57	0.39
	ST-DLR-AU	6.49	5.55	-	-
	ST-DLC-AU	-	-	54.54	0.42
	FS-MT-AU	**6.28**	**5.03**	54.54	0.44
	SP-MT-AU	6.46	5.42	**57.57**	**0.45**
	ST-DLR-COV	6.64	5.72	-	-
	ST-DLC-COV	-	-	51.51	0.36
	FS-MT-COV	**6.55**	**5.67**	**54.54**	**0.40**
	SP-MT-COV	6.59	5.71	54.54	0.37
	ST-DLR-FMT	6.91	5.89	-	-
	ST-DLC-FMT	-	-	51.51	0.34
	FS-MT-FMT	6.72	**5.77**	**54.54**	**0.36**
	SP-MT-FMT	**6.69**	5.79	51.51	0.34
	ST-DLR-TXT	4.90	3.99	-	-
	ST-DLC-TXT	-	-	60.60	0.45
	FS-MT-TXT	4.96	3.90	**66.66**	**0.53**
	SP-MT-TXT	**4.70**	**3.81**	60.61	0.42
Multimodal	ST-DLR-CombAtt	4.42	3.46	-	-
	MT-DLR-CombAtt	**4.24**	**3.29**	-	-
	ST-DLC-CombAtt	-	-	57.57	0.46
	MT-DLC-CombAtt	-	-	**60.61**	**0.48**
SOTA	$VFSC_{sem}$	4.46	3.34	-	-
	AW_{bhv}	5.54	4.73	-	-
	MMD	4.65	3.98	-	-

Results are still far from satisfactory

Combining regression and classification is beneficial

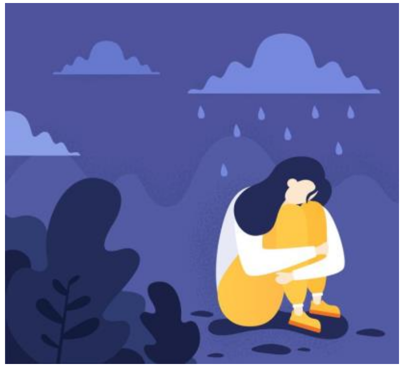
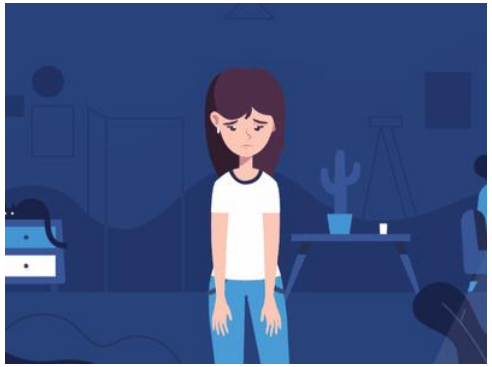
Language signal is strong

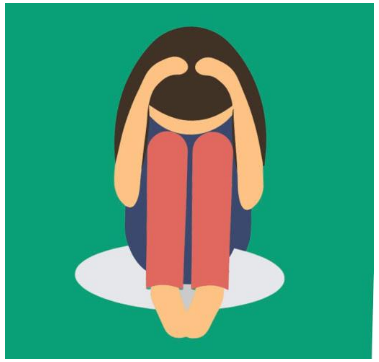
… But not for classification

Multimodality is beneficial …

Outline

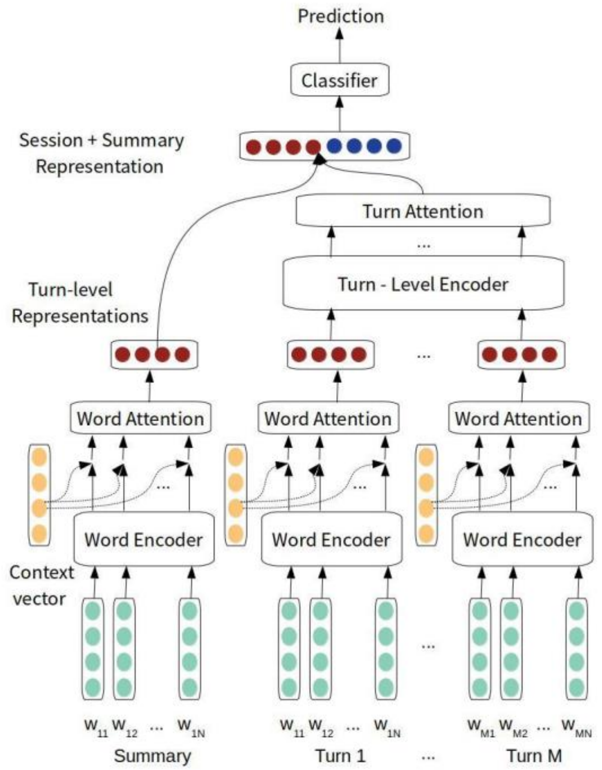
1. *Mental Health and Depression*
2. *6P Medicine*
3. *Interesting Iniatives*
4. **Computer-aided Diagnosis**
   i. *Multimodality*
   ii. *Emotionality*
   iii. *Gender-awareness*
   iv. *Dialogue structure*
   v. *Symptom-based diagnosis*
5. **A Favourable Research Environment**



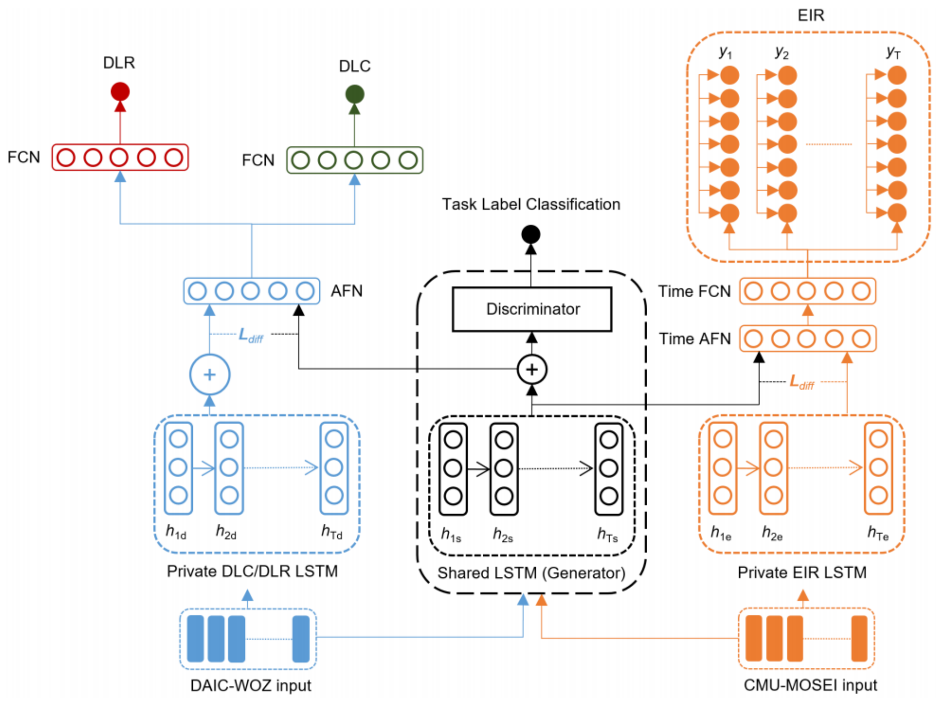
Emotional Language for the Estimation of PHQ-8

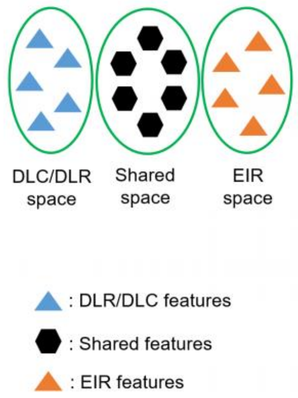
- Studies show that depression is a disorder of impaired emotion regulation.
- In particular, patients with major depression are often unable to control their emotional responses to negative situations, and overuse emotional expressions of sadness, disgust or fear.
- Emotion intensity can be evaluated on a Likert [0,3] scale for the six emotions of Ekman: happiness, sadness, anger, fear, disgust and surprise. But other models exist such as arousal and valence.
- In [Xezonaki et al., 2020], emotions are appended as an external context vector built from affective lexica (emotion, sentiment, valence).



Emotional Language for the Estimation of PHQ-8

- In [Qureshi et al., 2020], we hypothesize that the estimation of depression level can benefit from the concurrent learning of emotion intensity.
- The CMU-MOSEI dataset comprises 3,228 videos from 1,000 different speakers over 250 topics. Videos were gathered from an online video platform, where users emit their opinions in the form of monologues.





Emotional Language for the Estimation of PHQ-8

Not satisfactory for classification

Interesting results for regression, although with small improvements

Models	Evaluation Metrics													
	DLC							DLR						EIR
	Acc.	F1	MCC	RMSE	MAE	Ov.	Un.	RMSE	MAE	R^2	SM.	$\overline{Ov.}$	$\overline{Un.}$	$\overline{MSE}$
Baselines without Emotion Intensity Regression														
ST. DLC	60.61	0.54	0.38	1.31	0.75	3.03	36.36	-	-	-	-	-	-	-
ST. DLR	-	-	-	-	-	-	-	4.90	3.99	0.46	0.97	3.21	5.18	-
ST. EIR	-	-	-	-	-	-	-	-	-	-	-	-	-	7.15
FS MT. DLC+DLR	**66.66**	**0.62**	**0.49**	1.23	0.66	3.03	30.31	4.96	3.89	0.44	0.98	2.81	5.19	-
SP MT. DLC+DLR	60.61	0.51	0.39	1.26	0.72	0.00	39.39	4.70	3.81	0.50	0.99	3.39	4.32	-
Multi-task Results with Emotion Intensity Regression														
FS MT. DLC+EIR	60.61	0.51	0.42	1.58	0.90	0.00	39.39	-	-	-		-	-	6.98
SP MT. DLC+EIR	57.57	0.50	0.35	1.27	0.76	6.07	36.36	-	-	-	-	-	-	7.05
ASP MT. DLC+EIR	60.61	0.54	0.38	1.26	0.73	9.09	30.30	-	-	-	-	-	-	7.19
FS MT. DLR+EIR	-	-	-	-	-	-	-	4.60	3.74	0.52	0.99	3.16	4.63	6.88
SP MT. DLR+EIR	-	-	-	-	-	-	-	**4.51**	3.89	**0.54**	**0.94**	3.91	3.85	**6.82**
ASP MT. DLR+EIR	-	-	-	-	-	-	-	4.72	3.96	0.50	**0.94**	3.80	4.15	7.08
FS MT. DLC+DLR+EIR	57.57	0.46	0.38	1.36	0.82	3.04	39.39	4.83	4.03	0.47	0.97	3.13	5.11	6.96
SP MT. DLC+DLR+EIR	**63.64**	0.58	**0.48**	**0.94**	**0.51**	24.24	12.12	4.56	3.79	0.53	0.97	3.20	4.59	7.02
ASP MT. DLC+DLR+EIR	60.61	**0.60**	0.42	1.14	0.64	12.12	27.27	4.61	**3.69**	0.52	0.95	2.87	4.81	7.11

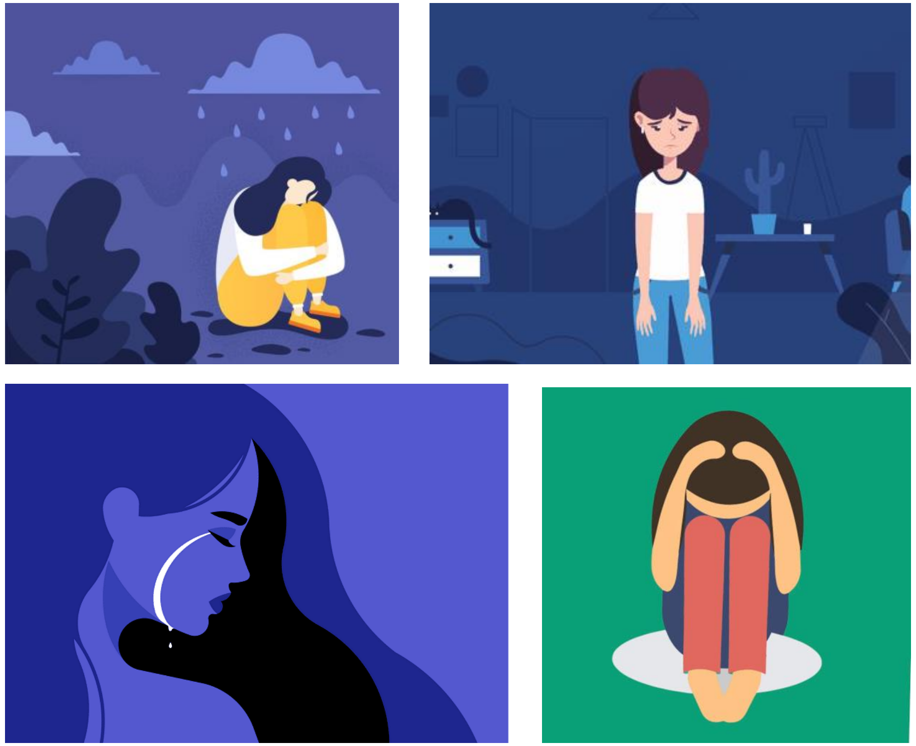
High standard deviation per class

Strong under-evaluation for the moderately severe class

Models	Evaluation Metrics								
	DLC					DLR			
	Acc.	RMSE	MAE	Ov.	Un.	RMSE	MAE	$\overline{Ov.}$	$\overline{Un.}$
	Best for DLC without EIR: FS MT. DLC+DLR					**Best for DLR without EIR: SP MT. DLC+DLR**			
None-minimal	100	0.00	0.00	0	-	3.97	3.22	3.51	1.14
Mild	40	1.10	0.80	20.00	40	3.80	3.11	3.82	2.05
Moderate	40	1.34	1.00	0.00	60	4.04	3.50	0.90	3.50
Moderately severe	33.33	2.27	1.83	0.00	66.67	6.78	5.75	0.47	6.81
Severe	0	2.00	2.00	-	100	6.81	6.81	0.00	6.81
	Best for DLC+EIR: ASP MT. DLC+EIR					**Best for DLR+EIR: SP MT. DLR+EIR**			
None-minimal	100	0.00	0.00	0	-	4.28	3.85	4.05	0.74
Mild	20	1.18	1.00	40	40	3.51	3.07	3.56	2.32
Moderate	20	1.61	1.40	20	60	2.94	2.60	0.00	2.60
Moderately severe	33.33	2.16	1.67	0	66.67	6.70	6.05	2.77	6.71
Severe	0	2.00	2.00	-	100	2.03	2.03	0.00	2.03
	Best for DLC+DLR+EIR: SP MT. DLC+DLR+EIR								
None-minimal	93.75	0.50	0.13	6.25	-	3.42	2.89	2.97	1.79
Mild	0	1.00	1.00	60	40	3.78	3.49	3.89	2.88
Moderate	80	0.89	0.40	0	20	3.84	3.37	0.00	3.37
Moderately severe	33.33	1.41	1.00	0	66.67	7.54	6.78	4.67	7.21
Severe	0	2.00	2.00	-	100	3.85	3.85	0.00	3.85

Outline

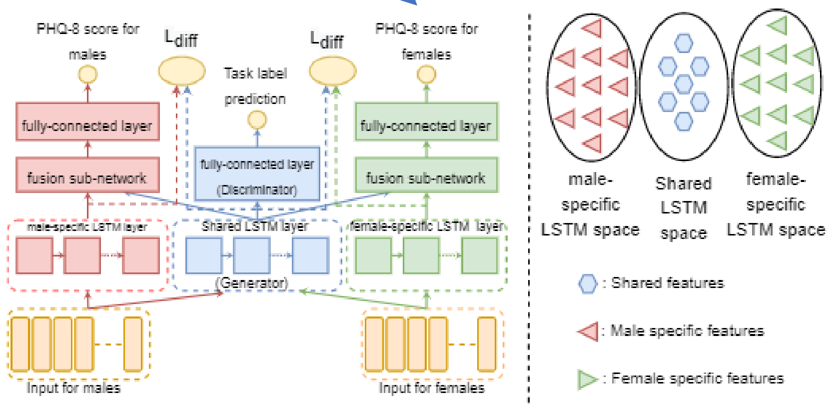
1. *Mental Health and Depression*
2. *6P Medicine*
3. *Interesting Iniatives*
4. **Computer-aided Diagnosis**
   i. *Multimodality*
   ii. *Emotionality*
   iii. *Gender-awareness*
   iv. *Dialogue structure*
   v. *Symptom-based diagnosis*
5. **A Favourable Research Environment**



Gender-awareness for the Estimation of PHQ-8

- [Joan and Kaite, 2015] reviewed several works in psychological research on the difference in gender in depression.
  - They state that by the middle of adolescence, females are about twice as likely to be diagnosed with depression and exhibit twice as many depressive symptoms as males, and this trend may continue till they are at least 55 years old.
- However, very few works have been proposed on how depression is dependent on gender.
- In [Qureshi et al., 2021], we propose to study gender-aware models in multimodal settings.

- Depression estimation without gender information (Gen_{less})
- Depression estimation with concatenated gender information (Gen_{concat})
- Multitask prediction of depression level and gender (Gen_{pred})
- Multitask prediction of depression level in males and females separately, using shared-private multitask network [35] (Gen_{SP})
- Multitask prediction of depression level in males and females separately, using adversarial shared-private multitask network [35] (Gen_{ASP})



Gender-awareness for the Estimation of PHQ-8

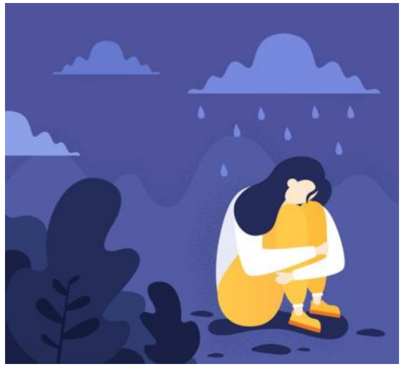
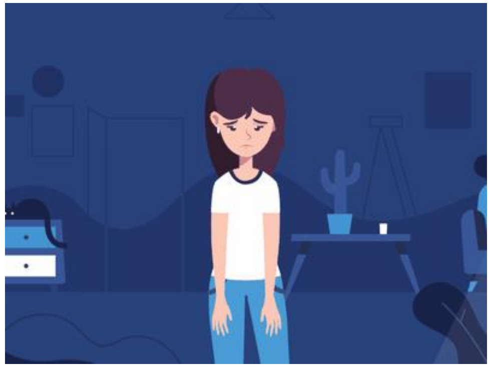
Models	Evaluation Metrics									
	Gen_{less}		Gen_{concat}		Gen_{pred}		Gen_{SP}		Gen_{ASP}	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
COVAREP	44.98	5.32	44.59	5.16	**43.05**	5.14	43.70	**5.11**	44.13	5.14
Formant	43.11	5.48	42.21	5.50	42.29	5.54	42.53	5.56	**41.96**	**5.19**
Facial action units	42.32	5.51	41.97	**5.13**	**41.06**	5.47	41.90	5.15	41.97	5.21
Eye gaze	47.26	5.57	47.04	5.62	46.05	5.75	48.01	5.72	**44.41**	**5.23**
Facial landmarks	52.82	6.21	50.72	6.06	52.45	5.93	47.16	5.87	**45.13**	**5.51**
Head pose	48.99	5.78	47.31	5.74	46.92	5.76	46.56	5.54	**44.29**	**5.40**
Text	23.82	**3.78**	23.28	3.87	**23.12**	3.87	24.12	4.10	24.02	4.09
Multimodal	24.12	3.74	**20.06**	3.50	20.56	3.50	21.01	3.51	22.25	**3.49**

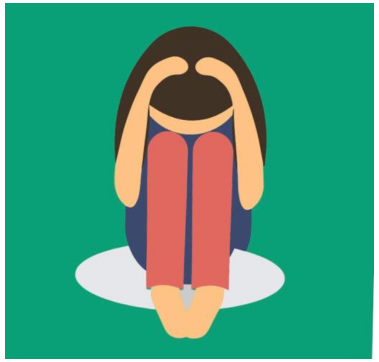
Strong indicator for the visual signal

Text is not so sensitive to gender!

Gender-awareness is important

Outline



GREYC Electronics and Computer Science Laboratory

Analysis of Structured Interviews

GREYC
Electronics and Computer Science Laboratory

- First observation: most of the related works have been dealing with the interview on a line by line basis; the hypothesis being that sentence representation is the correct one.
- Second observation: some of the related works only deal with the patient information; the hypothesis being that only the patient information is important for the diagnosis.
- Our hypothesis is that better diagnosis can be established if the correct level of language analysis is performed.

Sentence level

Question/Answer pair level

Topic/Chunk level

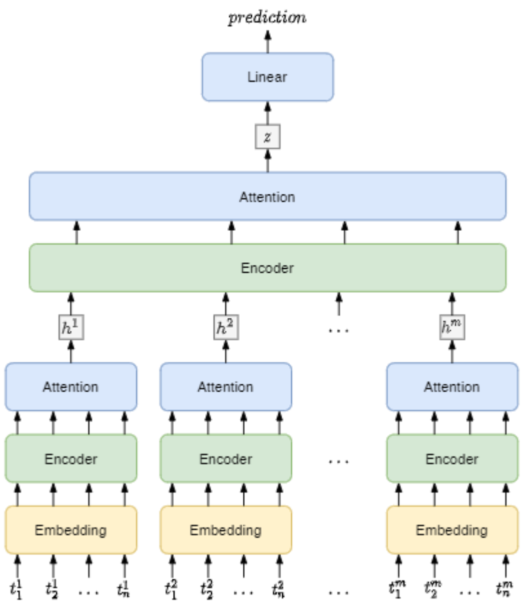
Information is conveyed by the therapist

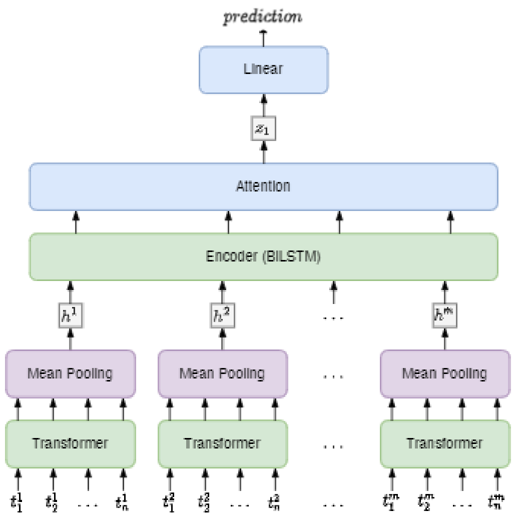
	Opening Rapport Building Phase
Ellie	*What are some things you really like about LA?* (top level question)
User	*I love the weather, I love the palm trees, I love the beaches, there's a lot to do here.*

	Diagnostic Phase
Ellie	*Have you noticed any changes in your behavior or thoughts lately?* (top level question)
User	*Yes.*
Ellie	*Can you tell me about that?* (continuation prompt)
User	*I'm having a lot more nightmares now uh can't sleep have haven't really been eating uh trying to eat... I have to force down food um just feeling like an emotional wreck.*
Ellie	*I'm sorry to hear that.* (empathy response)
Ellie	*What are you like when you don't sleep well?* (follow-up question)
User	*Irritable, emotional, it just adds to my overall stress um [long pause]*
Ellie	*What...* (Ellie speaks after the participant's long pause)
User	*Can't concentrate uh I uh...* (the participant starts speaking while Ellie is speaking)
Ellie	*I'm sorry please continue.* (Ellie realizes that she has interrupted the participant and apologizes)

Segmentation Level Analysis

- We propose to segment interviews into three different linguistic levels: sentences, question/answer pairs and semantic chunks.
- For that purpose, the DAIC-WOZ has been manually annotated at chunk level.
- To verify our hypothesis, we implement two different learning models: one on non-contextualized text embeddings [Xezonaki et al., 2020], and one with contextualized text embeddings.



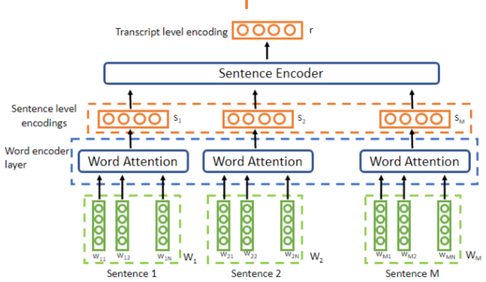


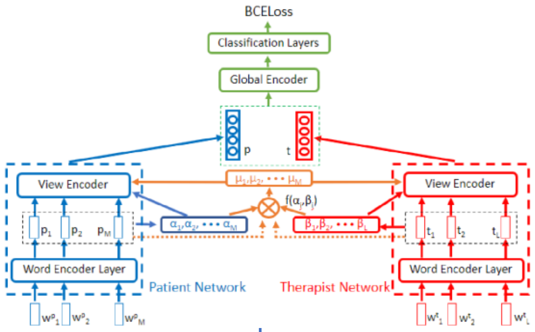
Model	Segmentation	F1	F1 (max)	MAE
GloVe+LSTM+LSTM	Lines	60.95 ± 8.02	72.63	5.606 ± 0.406
	Pairs	63.04 ± 7.84	79.55	5.678 ± 0.489
	Semantic	74.18 ± 9.05	87.59	5.302 ± 0.719
BERT+LSTM	Lines	91.38 ± 3.95	94.04	5.805 ± 0.440
	Pairs	87.71 ± 7.82	94.04	5.749 ± 0.638
	Semantic	74.69 ± 7.28	88.19	5.394 ± 0.507

Table: Results on the validation set. Each model was run 20 times with different random initializations. **F1** represents a macro-average F1-score and is given as a mean ± standard deviation; **F1 (max)** shows the maximum F1-score among 20 runs.

Patient vs. Therapist Information

- [Xenozaki, 2020] showed that both patient and therapist information convey information, but do not take advantage of this fact.
- So, we propose a multiview model that tackles both patient and therapist texts individually and then fuses the information to get a single prediction.
- Three different attention levels are proposed: local attention (patient OR therapist), cross attention (patient -> therapist and therapist -> patient), global attention (patient AND therapist).

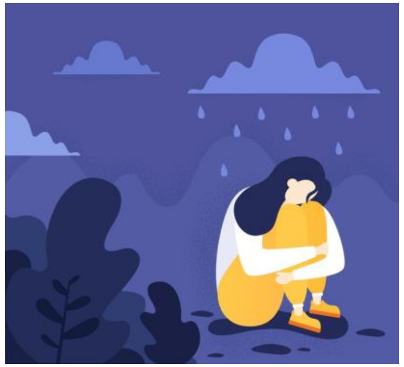
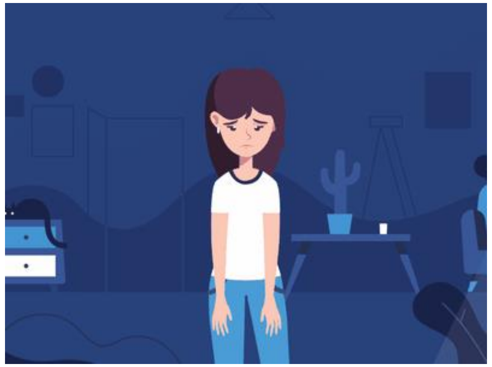


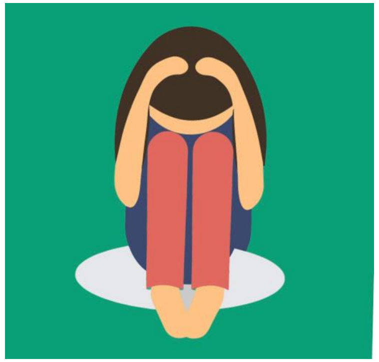


Architectures		macro F1 (Dev)	macro F1 Test	UAR (Dev)	UAR Test	Accuracy (Dev)	Accuracy Test	macro Precision (Dev)	macro Precision Test
Baseline	Patient	(0.6413)	0.6429	(0.6369)	0.6361	(0.6969)	**0.7608**	(0.6725)	0.6584
	Therapist	(0.8253)	0.5818	(0.8095)	0.5803	(0.8484)	0.6521	(0.8611)	0.6184
	Patient+Therapist	(0.7555)	0.6053	(0.7440)	0.6004	(0.7878)	0.6739	(0.7847)	0.6250
MV-Intra-Attention	View-Global Attention	(0.6944)	0.6811	(0.6845)	0.6674	(0.7575)	0.7391	(0.7870)	0.7252
	Global Attention	(0.6857)	0.7116	(0.6785)	0.7075	(0.7272)	0.7173	(0.7083)	0.6887
	View Attention	(0.6944)	0.6919	(0.6845)	0.6919	(0.7575)	0.6739	(0.7870)	0.6919
MV-Inter-Attention	Mean	(0.6857)	**0.7319**	(0.6785)	**0.7232**	(0.7272)	0.7173	(0.7083)	**0.7450**
	Learnable	(0.6434)	0.6043	(0.6428)	0.6093	(0.7272)	0.4782	(0.7571)	0.6020
	Max	(0.6616)	0.5801	(0.6845)	0.5982	(0.6666)	0.6304	(0.6709)	0.5846
	Patient	(0.5460)	0.5719	(0.5476)	0.5736	(0.6060)	0.6956	(0.5555)	0.5709
	Therapist	(0.7664)	0.5710	(0.7619)	0.5691	(0.7878)	0.6304	(0.7727)	0.5759

Table 2: Overall results over the DAIC-WOZ dataset. UAR stands for Unweighted Average Recall. We provide results for both development and test sets. The best model is chosen based on macro F1 over the development set.

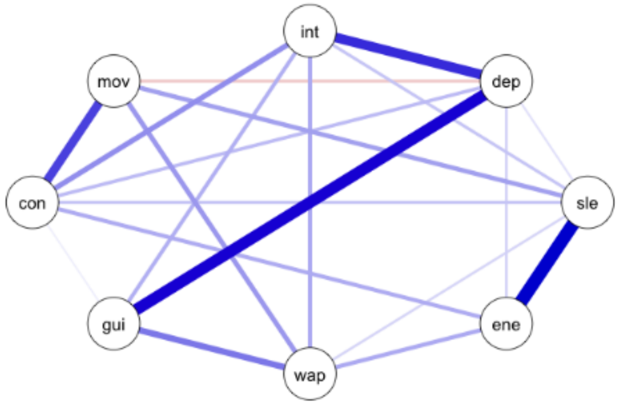
Outline



Symptom-based Analysis

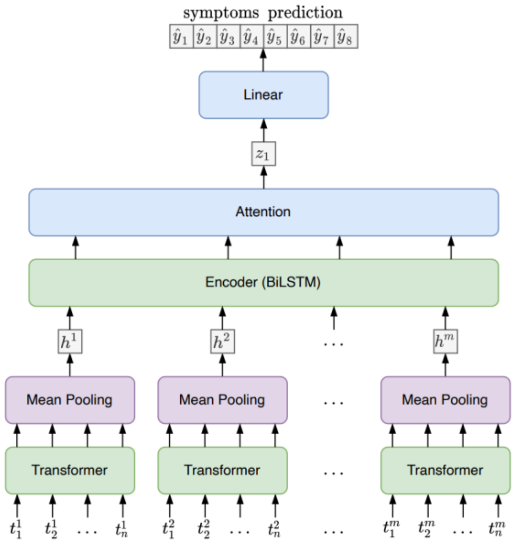
- Most related works have been tackling depression level estimation as a simple task (depressed or non depressed). More advanced models have been trying to predict the PHQ-8 score (between 0 and 24) directly or propose to solve the intermediate 5-class problem (none-minimal, mild, moderate, moderately severe, severe depression).
- In Psychiatry, there is a shift towards richer representations of psychiatric syndromes that can take into account the dimensional and heterogeneous nature of the clinical pictures of the same psychiatric diagnosis. One particular approach that is gaining attention concerns symptom network analysis.
- We develop similar models as previously to acknowledge if they can handle the prediction of individual symptom values, where each of the 8 symptoms is a value between 0 and 3.

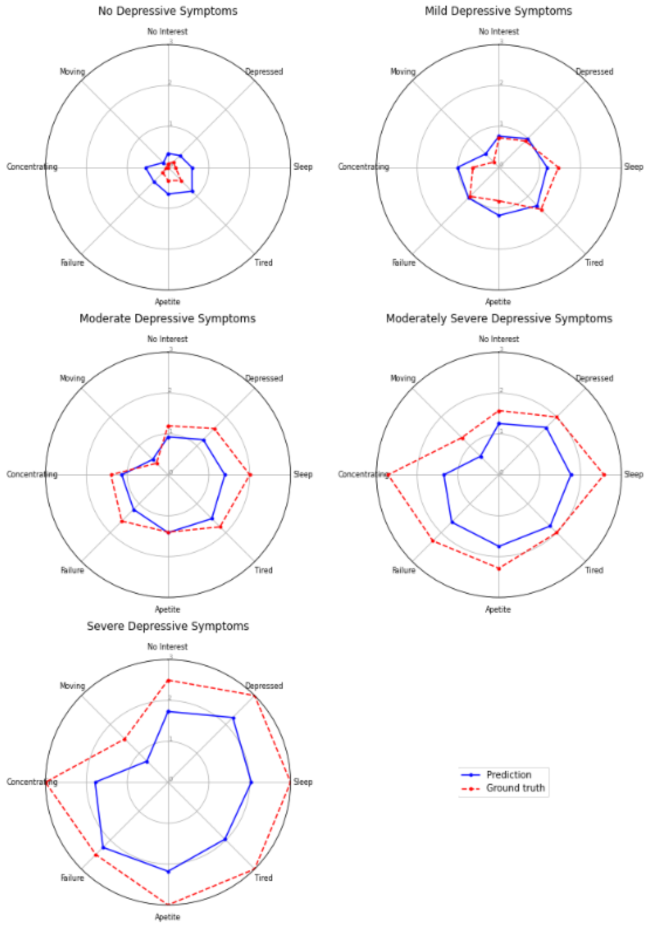


Symptom-based Analysis

GREYC
Electronics and Computer Science Laboratory

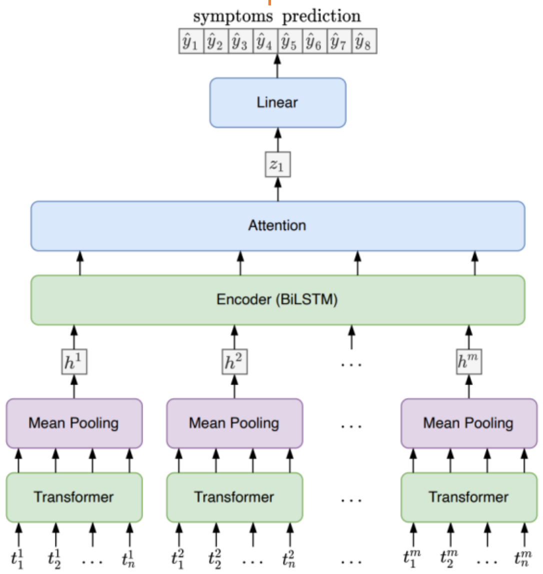
- In order to better understand results, we present a radar plot analysis that shows that adequate behavior of the model is obtained.





Symptom-based Analysis

- We evaluate the impact of categorical diagnosis based on symptoms prediction.

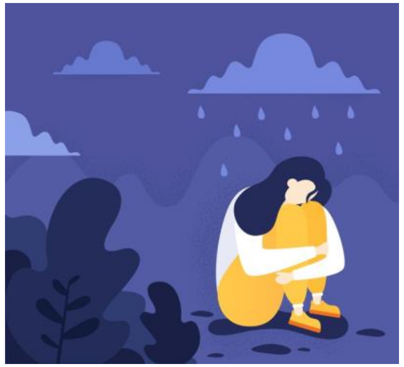
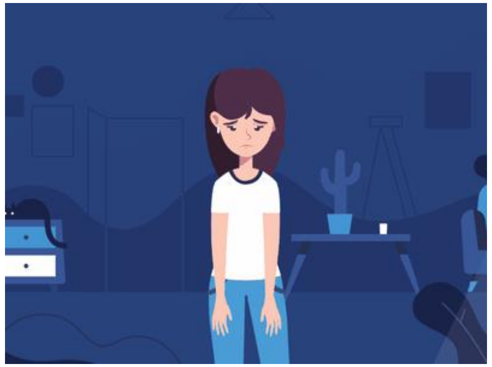


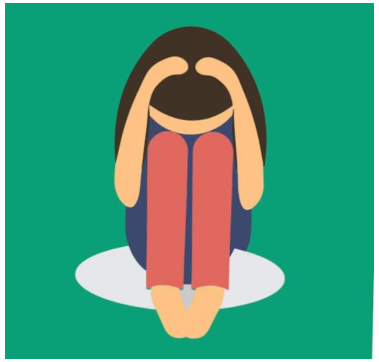
Model	$miF_1 \pm\sigma$	$maF_1 \pm\sigma$	miMAE $\pm\sigma$	maMAE $\pm\sigma$	$miF1$-5c $\pm\sigma$	$maF1$-5c $\pm\sigma$
Binary Diagnosis (BD)	71.91 ± 1.59	70.08 ± 1.03	-	-	-	-
5-Class Diagnosis (5CD)	71.06 ± 2.55	68.30 ± 2.39	-	-	**46.81** ± 2.33	**27.03** ± 2.46
PHQ-8 Score Diagnosis (PSD)	68.09 ± 1.90	58.44 ± 2.39	5.03 ± 0.09	5.69 ± 0.12	28.94 ± 2.89	13.49 ± 1.43
Symptom-based Diagnosis (SD)	76.60 ± 2.33	73.87 ± 2.48	3.78 ± 0.13	4.19 ± 0.13	42.55 ± 1.35	27.00 ± 1.93
SOTA results based on the text modality only						
HCAN [7] (2019)	-	†63.00	-	-	-	-
HAN+L [8] (2020)	-	†70.00	-	-	-	-
ASP MT. DLC+DLR+EIR [25] (2020)	-	-	3.69	-	60.00	-
HCAG-T [23] (2021)	-	†77.00	†3.73	-	-	-
SGNN [27] (2022)	-	-	†3.76	-	-	-
SOTA results based on the multiple modalities						
SVM:m-M&S [9] (2021)	-	67.00	3.98	-	-	-
Gen$_{ASP}$ [26] (2021)	-	-	3.49	-	-	-
MFCC-AU [31] (2021)	-	66.50	-	-	-	-
HCAG-A+T [23] (2021)	-	†92.00	†2.94	-	-	-
BLSTM [28] (2022)	-	-	-	-	‡95.80	-

Table 2 Experimental and state-of-the-art results over the test set of the DAIC-WOZ. Models are run five times with different seed values for BD, 5CD, PSD and SD, so that average values with standard deviation are presented. Note that $miF1$-5c (resp. $maF1$-5c) stand for the 5-class micro-averaged F1-score (resp. macro-averaged F1-score). Note that "†" indicates that results are given for the best configuration and not based on average performance. Note also that "‡" indicates that results are given for a balanced test set and not the original test set provided with the DAIC-WOZ.

Outline

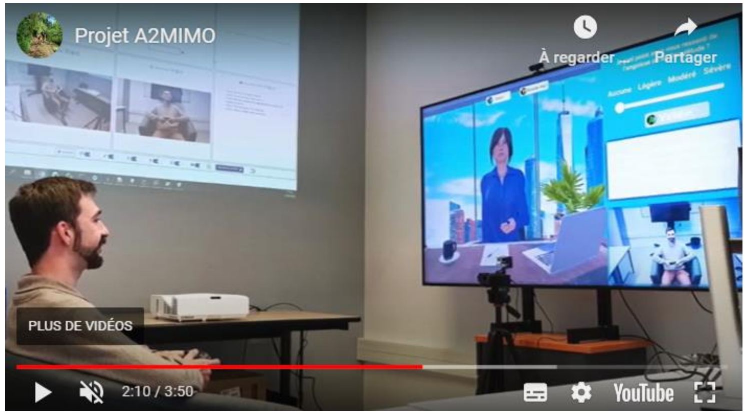
1. *Mental Health and Depression*
2. *6P Medicine*
3. *Interesting Iniatives*
4. *Computer-aided Diagnosis*
   i. *Multimodality*
   ii. *Emotionality*
   iii. *Gender-awareness*
   iv. *Dialogue structure*
   v. *Symptom-based diagnosis*
5. **A Favourable Research Environment**

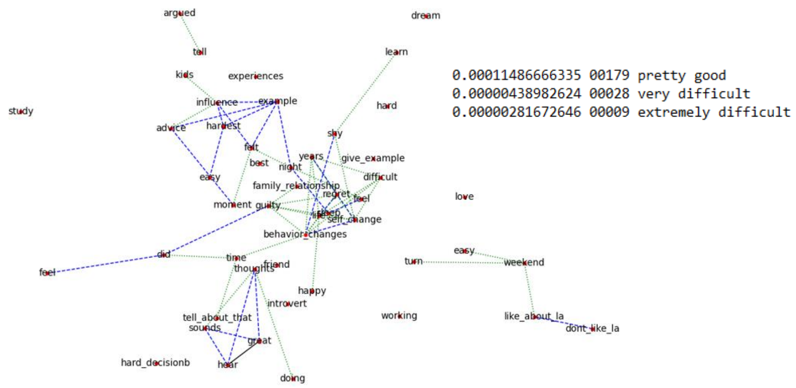





On-going Research @ CAEN

- Navneet Agarwal
  - Graph-based representation and learning.
  - Psychiatrists into the loop.
- Kirill Milintsevich
  - External knowledge introduction.
  - Dataset quality assessment.
- Soumaya Sabry
  - Embodied conversational agents for early detection.
  - Therapeutic alliance.





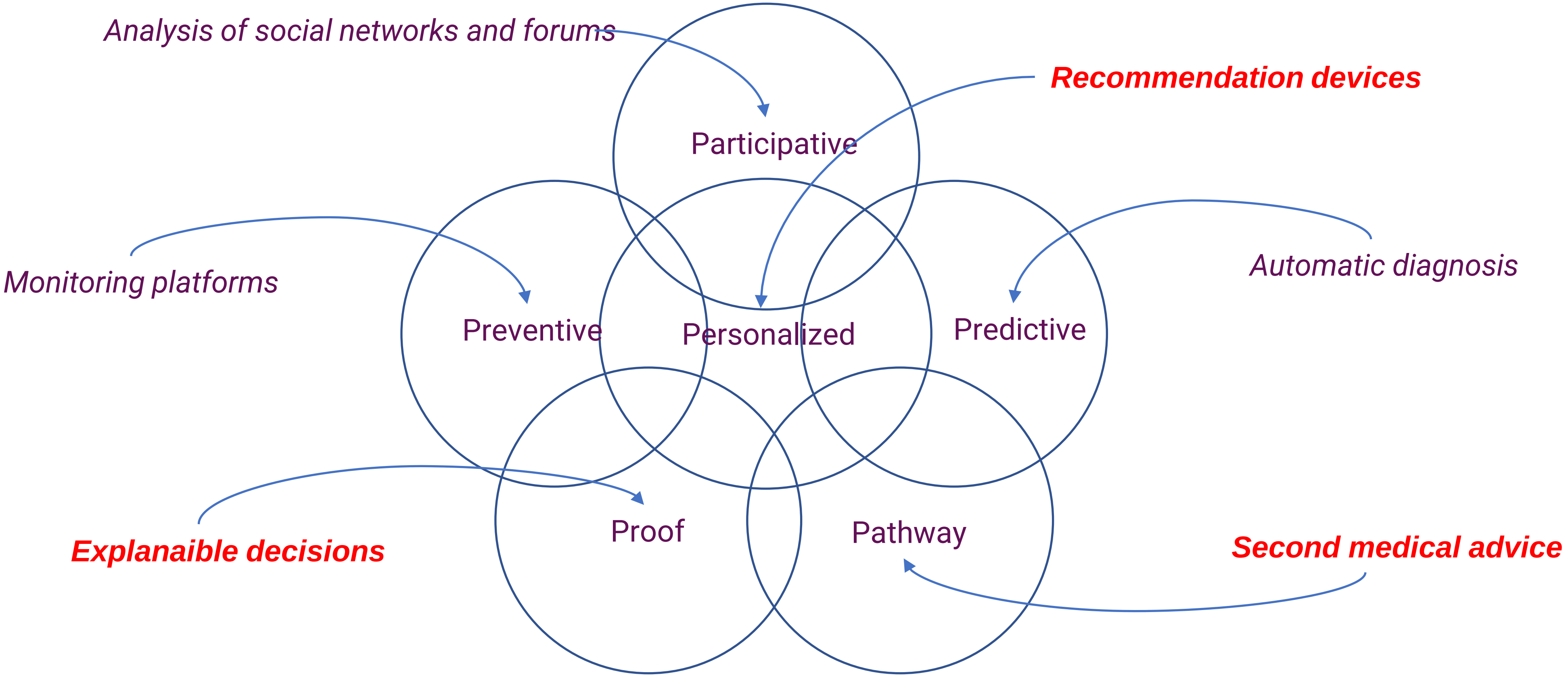
Other Running Projects

- Prediction of suicidal recidivism from phone conversations.
  - Pr. Françoise CHASTANG and Dr. Pierre GERARD.
  - Project VigilanS.
  - CHU Estran.
- Automatic level estimation of schizophrenia from emergency interviews.
  - Pr. Christophe LEMEY and Pr. Sonia DOLLFUS.
  - Project ASESID.
  - CHU Brest.

Structuring Mental Health in Normandy

- FHU A2M2P - Improving the prognosis of addictive and mental disorders through personalized medicine (Améliorer le pronostic des troubles Addictifs et Mentaux par une Médecine Personnalisée).
    - 5-year project including 11 research laboratories (CNRS, INSERM, EA), 4 hospitals (Amiens, Caen, Rouen), patient and family associations, public health services (ARS).
    - Jointly studying mental disorders and drug addiction.
- Department of Mental Health and Digital Sciences at BB@C (Blood and Brain GIS).
    - Gathering worldwide specialists in AI and Mental Health inside the same research structure.
    - Initiative of the Pole TES competitiveness cluster and the agglomeration of Caen.
- Second Workshop on Mental Health and Artificial Intelligence @ Caen
    - January 29th -30th, 2024.
    - https://mentalai.ubi.pt/symposium

Un(less)explored Areas



THANK YOU FOR YOUR ATTENTION

Freel free to ask caring questions ;)

Gaël DIAS @ Sorbonne

joint work with
Navneet AGARWAL, Mohammed HASANUZZAMAN, Arbaaz QURESHI, Kirill MILINTSEVICH, Valentin RENIER, Soumaya SABRY, Sriparna SAHA, Kairit SIRTS, and more to come ;)

University of Caen Normandie - CNRS - GREYC UMR 6072 Research Laboratory
gael.dias@unicaen.fr

GREYC Electronics and Computer Science Laboratory

Normandie Université UNICAEN ENSICAEN ÉCOLE PUBLIQUE D'INGÉNIEURS CENTRE DE RECHERCHE cnrs