

Ce document est copiable et distribuable librement et gratuitement à la condition expresse que son contenu ne soit modifié en aucune façon, et en particulier que le nom de son auteur et de son institution d'origine continuent à y figurer, de même que le présent texte.



**Attribution-NonCommercial-NoDerivatives 4.0
International**

<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

VOCAL AND SPEECH BIOMARKERS OF PSYCHIATRIC DISORDERS

Vincent P. MARTIN

February 1, 2023

LaBRI

université
de **BORDEAUX**

 **SANPSY**
BORDEAUX
neurocampus

HELLO!

I am Vincent P. MARTIN

- ▶ **Ph.D. (2022)** « Voice biomarkers of sleepiness », *Université de Bordeaux*
J.L. Rouas (LaBRI) & P. Philip (SANPSY/CHU)
- ▶ **Eng. Degree (2018)**
Ecole Nationale Supérieure de l'Electronique et de ses Applications (ENSEA)
- ▶ **DIU Philosophy of psychiatry (2021)**
Université de Bordeaux



vincent.martin@labri.fr



@V_P_Martin



Vincent-P-Martin

1.

The need for **objective** diagnosis?

A NEED FOR OBJECTIVE DIAGNOSIS DEVICES ?

“There is an **urgency** to **objectively diagnose**, monitor over time, and provide evidence-based interventions for individuals with mental illnesses”

[Low et al. 2020](#)

“Gold-standard diagnostic and assessment tools for depression and suicidality remain rooted, almost exclusively, on the **opinion of individual clinicians** risking a range of **subjective biases**. [...] Currently there is no **objective measure**, with **clinical utility**, for either depression or suicidality”

[Cummins et al. 2015](#)

A NEED FOR OBJECTIVE DIAGNOSIS DEVICES ?

Aboraya 2007

- ▶ 28 healthcare professionals
- ▶ Most of them (87%) : their diagnosis is not trustworthy
- ▶ Why ?
 - ▶ Factors linked to clinicians (education, biases, style) : **63.5%**
 - ▶ Patients characteristics : **21.6%**
 - ▶ Disease definitions **14.9%**

**1rst factor for diagnosis
= clinician**

A NEED FOR OBJECTIVE DIAGNOSIS DEVICES ?

Kendell 1971

Table 6.—Diagnoses Given to Patient F

	American Psychiatrists (N = 133)	British Psychiatrists (N = 194)
Schizophrenia	92 (69%)	4 (2%)
Simple	0	1
Catatonic	1	0
Paranoid	27	1
Latent	8	0
Residual	3	0
Schizo-affective	33	1
Unspecified	20	1
Personality Disorder	10 (8%)	146 (75%)
Paranoid	1	2
Affective (cyclothymic)	1	8
Explosive	0	2
Hysterical	4	105
Asthenic	0	2
Antisocial	1	8
Unspecified	3	19
Affective Psychosis	10 (8%)	7 (4%)
Neurosis	19 (14%)	37 (19%)
Alcoholism or Drug Dependence	2	0

⇒ We need **objective** diagnosis

2.

Voice biomarkers of psychiatric disorders

State of the art

STATE OF THE ART

- ▶ [Low et al. 2020](#),
« Automated assessment of
psychiatric disorders using
speech: A systematic review »,
*Laryngoscope Investigative
Otolaryngology*



STATE OF THE ART Low et al. 2020

METHOD

- ▶ Google Scholar
- ▶ 2009-2019

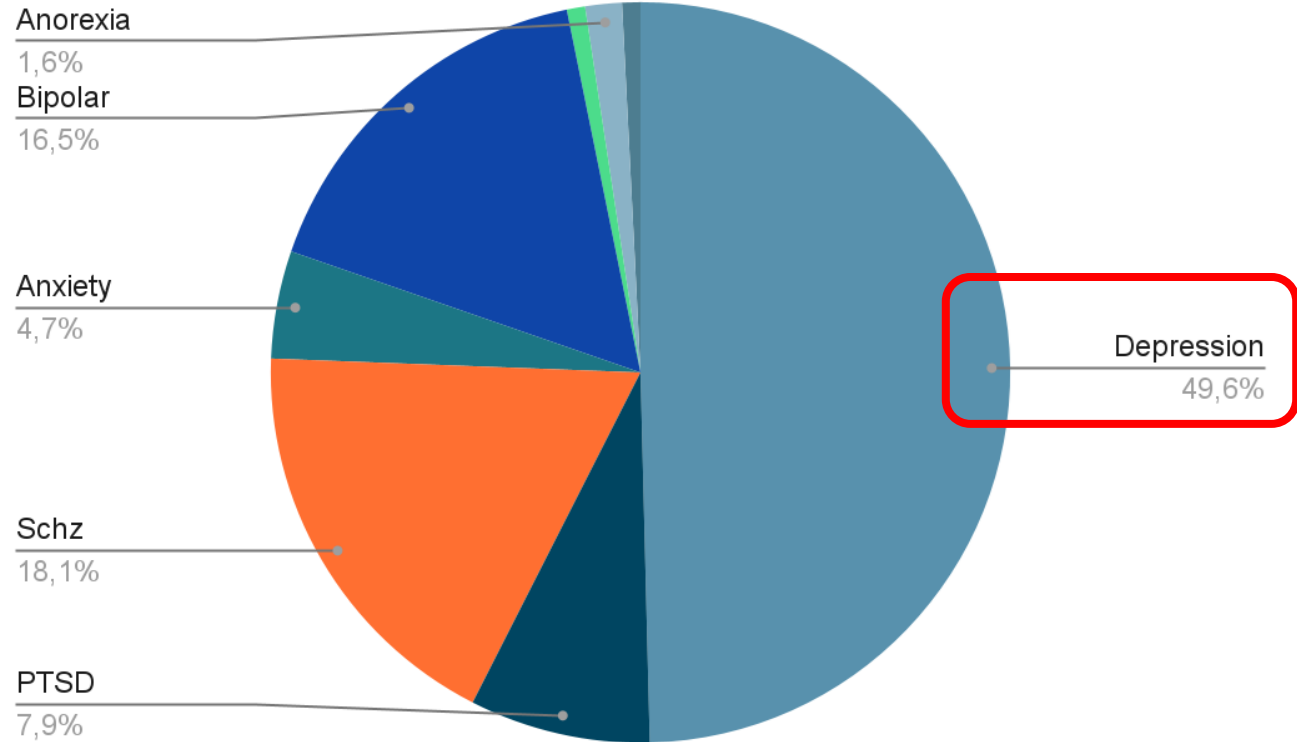
- ▶ **127** studies



Supplementary data available online!

STATE OF THE ART Low et al. 2020

RESULTS



***Pb : not used in **real**
clinical settings***

NOT USED IN REAL CLINICAL SETTINGS


WHY ?

- ▶ Performances ?
 - ▶ **80%** for bipolar disorders, **95%** for schizophrenia, **89.3%** for depression
 - ▶ High performances for decades (e.g. 75% for depression)



NOT USED IN REAL CLINICAL SETTINGS



WHY ?

- ▶ Performances ? 
- ▶ Databases size?
 - ▶ Ecological data collection
 - ◆ n=9920 ([Rutowski et al. 2022](#))
 - ◆ n=3580 ([Di et al. 2021](#))



NOT USED IN REAL CLINICAL SETTINGS




WHY ?

- ▶ Performances ? 
- ▶ Databases ? 
- ▶ Reglementary limitations ?



NOT USED IN REAL CLINICAL SETTINGS

WHY ?

- ▶ Performances ? 
- ▶ Databases ? 
- ▶ Reglementary limitations ? 
- ▶ Explanation and transparency ?
 - ▶ **TRUST**



NOT USED IN REAL CLINICAL SETTINGS

WHY ?

- ▶ Performances ? ❌
- ▶ Databases ? ❌
- ▶ Reglementary limitations ? ❌
- ▶ Explanation and transparency ? ❌



THERAPEUTIC RELATIONSHIP



- ▶ **Bourla et al.:** [*Bourla et al. 2018*](#)
 - ▶ 515 psychiatrists
 - ▶ 3 scenarios: biosensors comprising a connected wristband-based digital phenotype, ML-based blood test, ML-based magnetic resonance imaging (MRI).
 - ▶ 4 acceptability domains usefulness, usability, reliability, and risk

 - ▶ Overall acceptability=moderate.
 - ▶ **All systems = risky (410/515, 79.6%).**
 - ▶ Acceptability = strongly influenced by socioepidemiological variables (professional culture), such as gender, age, and theoretical approach.
 - ▶ **Worries = therapeutic relationship, data security, data storage, and privacy risk**

THERAPEUTIC RELATIONSHIP

Bourla et al. 2018

- ▶ Important for treatment adherence and therapeutic issues
- ▶ « You may have depression », « You have a probability of XX% of having schizophrenia », ...

3.

Diagnostic : what are we talking about?

STATE OF THE ART Low et al. 2020

RESULTS

Label

- ▶ Questionnaires (ex. PHQ9)



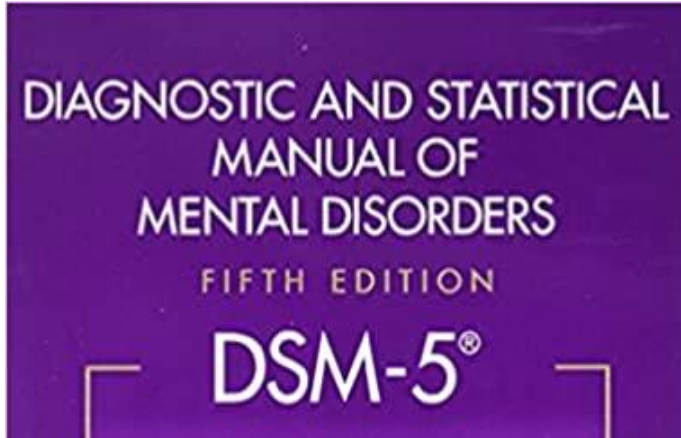
		Not at all	Several days	More than half the days	Nearly every day
1.	Little interest or pleasure in doing things	0	1	2	3
2.	Feeling down, depressed, or hopeless	0	1	2	3
3.	Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4.	Feeling tired or having little energy	0	1	2	3
5.	Poor appetite or overeating	0	1	2	3
6.	Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0	1	2	3
7.	Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8.	Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9.	Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3

STATE OF THE ART Low et al. 2020

RESULTS

Label

- ▶ Questionnaires (ex. PHQ9)
- ▶ Classification (e.g., DSM or ICD)



Major Depressive Disorder

Diagnostic Criteria

- A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

Note: Do not include symptoms that are clearly attributable to another medical condition.

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note:** In children and adolescents, can be irritable mood.)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (**Note:** In children, consider failure to make expected weight gain.)
4. Insomnia or hypersomnia nearly every day.
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

STATE OF THE ART (Low et al. 2020)

RESULTS

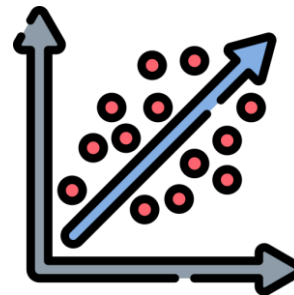
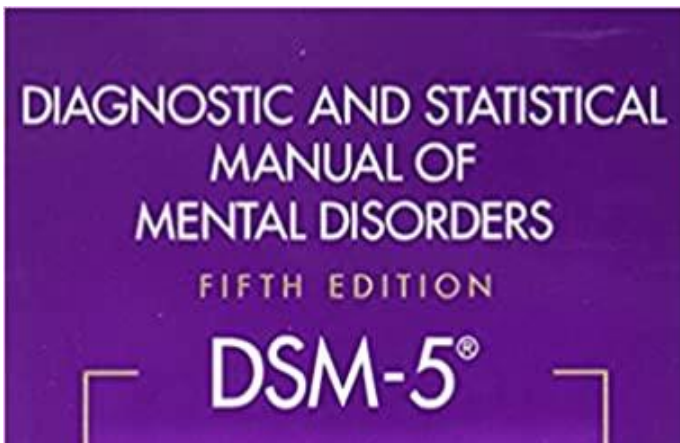


Label

- ▶ Questionnaires (ex. PHQ9)
- ▶ Classification (e.g., DSM or ICD)

Tasks

- ▶ **diagnostic:** binary classification
- ▶ **severity estimation:** regression with score



DIAGNOSIS : LIMITS

Questionnaires

- ▶ Not used by clinicians
- ▶ Validated on diagnosis criteria

Diagnosis criteria

- ▶ Culture (hikikomori)
- ▶ Time (versions of the DSM)
- ▶ **Heterogeneity**

Depression

- ▶ Number of semiological profiles
- ▶ $n = \binom{2}{1} \times \left(\binom{8}{4} + \binom{8}{5} + \dots + \binom{8}{8} \right)$
- ▶ = **326 unique profiles**
- ▶ **Eiko Fried: [STAR*D \(2015\)](#) : 1030 profiles on 3703 “depressive” patients**

Major Depressive Disorder

Diagnostic Criteria

- A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.
- Note:** Do not include symptoms that are clearly attributable to another medical condition.
1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note:** In children and adolescents, can be irritable mood.)
 2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
 3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (**Note:** In children, consider failure to make expected weight gain.)
 4. Insomnia or hypersomnia nearly every day.
 5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
 6. Fatigue or loss of energy nearly every day.
 7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
 8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
 9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

DIAGNOSIS : LIMITS

All pathologies [Newson 2021](#)

- ▶ **107349** patients
- ▶ **10** most prevalent disorders
- ▶ **47** symptoms

- ▶ Conclusion : « DSM-5 disorder criteria **do not separate individuals from random** when the complete mental health symptom profile of an individual is considered.»

So, diagnosis == useless ?



WHAT IS THE DIAGNOSIS USEFUL FOR ?

All pathologies

- ▶ Acknowledgement from **medical specialist** and from **society**
- ▶ **10** most prevalent disorders
- ▶ **47** symptoms

- ▶ Conclusion :
 - « DSM-5 disorder criteria **do not separate individuals from random** when the complete mental health symptom profile of an individual is considered.»

WHAT IS THE DIAGNOSIS USEFUL FOR ?

“the main aim of the psychiatric science **is not classification** as an end in itself but rather **identification of causes** and **interventions**”

Keneth Kendler, 2012

“[...] **classification in itself is** less important than often supposed to be, and **less important than other tasks.**”

Derek Bolton, 2012

« [...] one of its most important goal is to **facilitate communication among clinicians, researchers, administrators and patients** [...] by establishing a common language.”

Derek Bolton, 2012

- + **prognostic**
- + **differential** diagnosis

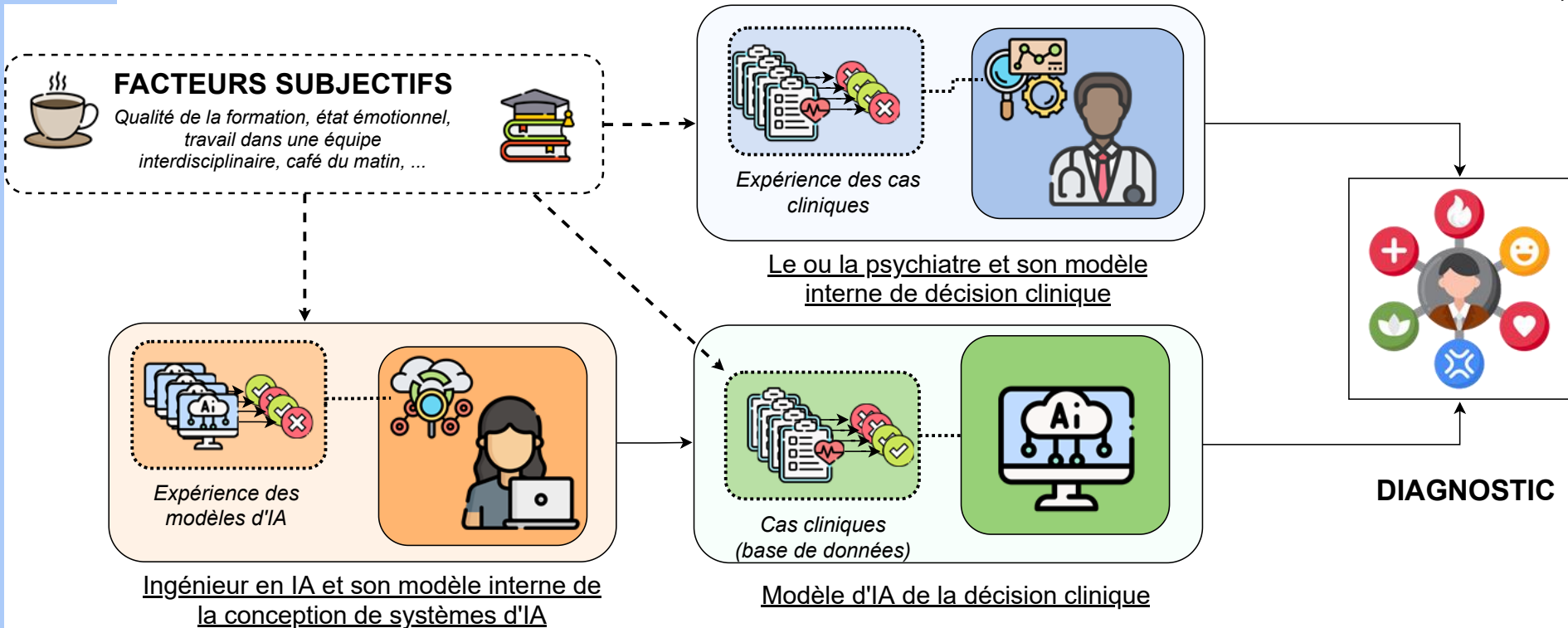
*What is an
« objective diagnosis » ?*



How Does Comparison With Artificial Intelligence Shed Light on the Way Clinicians Reason? A Cross-Talk Perspective

Vincent P. Martin^{1,2}, Jean-Luc Rouas¹, Pierre Philip^{2,3}, Pierre Fourneret⁴, Jean-Arthur Micoulaud-Franchi^{2,3} and Christophe Gauld^{4,5}*





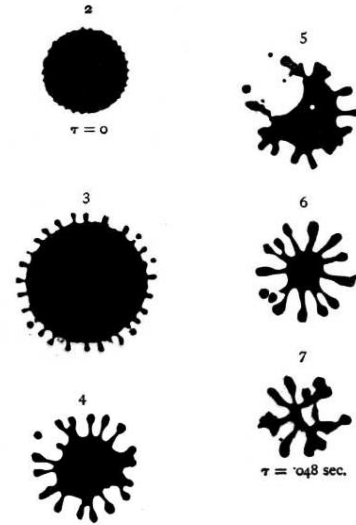
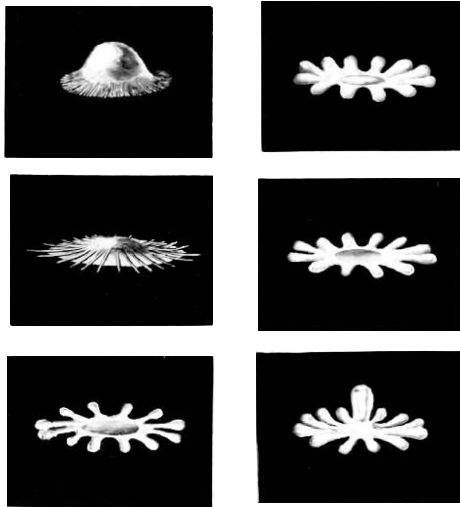
OBJECTIVITY : A GOLD-STANDARD?

36

THE SPLASH OF A DROP, Pr. Worthington, *Royal Institution of Great Britain*, May 18, 1894

- ▶ Mercury drop on glass
- ▶ Light at constant delay

Which one represents **reality** more accurately?

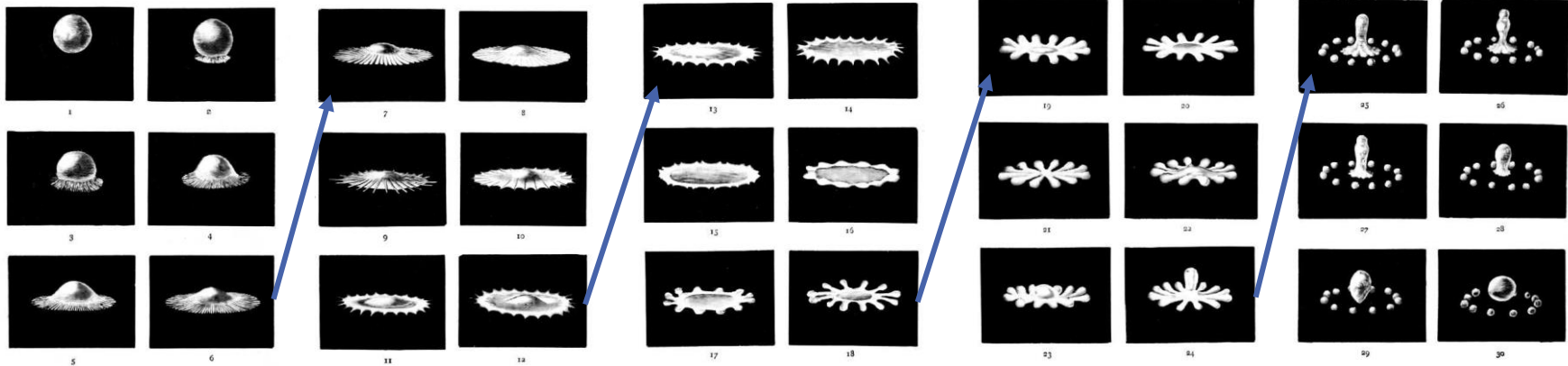


OBJECTIVITY : A GOLD-STANDARD?

37

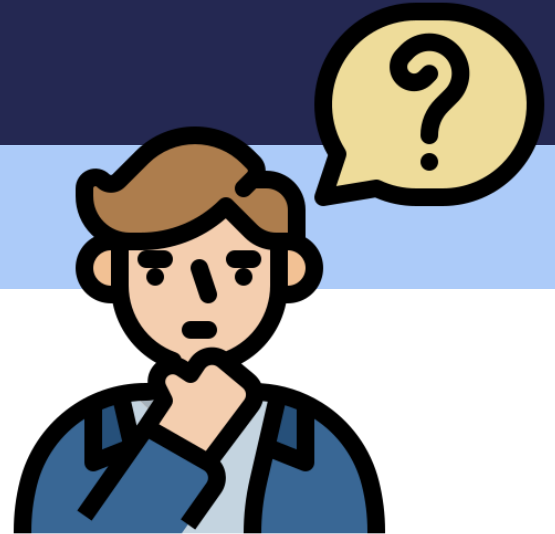
THE SPLASH OF A DROP, Pr. Worthington, *Royal Institution of Great Britain*, May 18, 1894

- ▶ Mercury drop on glass
- ▶ Light at constant delay



4.

What could be useful for clinical practice ?



Symptoms



SYMPTOMS



Diagnosis	Symptoms
Time dependent <i>e.g. DSM IV, DSM 5, ...</i>	Stable through time
Cultural dependent <i>e.g. Hikikomori</i>	Independent from culture
Heterogeneous	Homogeneous
Symptoms → Syndromes → Diagnosis	
-	
-	



SYMPTOMS



Diagnosis	Symptoms
Time dependent <i>e.g. DSM IV, DSM 5, ...</i>	Stable through time
Cultural dependent <i>e.g. Hikikomori</i>	Independent from culture
Heterogeneous	Homogeneous
Symptoms → Syndromes → Diagnosis	
-	Mechanistic explanation
-	



SYMPTOMS



Diagnosis	Symptoms
Time dependent <i>e.g. DSM IV, DSM 5, ...</i>	Stable through time
Cultural dependent <i>e.g. Hikikomori</i>	Independent from culture
Heterogeneous	Homogeneous
Symptoms → Syndromes → Diagnosis	
-	Mechanistic explanation
-	Necessary for diff. diag and prog.



NUMERIC SYMPTOMS

“subjects have no need to be equipped with multiple sensors or even be burdened by **invasive devices** (e.g., endoscopy) [...] Additionally, CA can make it feasible to collect data from subjects via mobile devices (e.g., a smartphone), which can provide the subjects **24×7 monitoring service.**”

Qian et al. 2020

- + **no judgment**
- + **easy to access** (smartphones)
- + **no biases** (patients and clinicians)



Yes.

NUMERIC SYMPTOMS

45

Anchoring bias	Tendency to focus on a first impression or on the first information received to form an opinion about a number, a person, an event ... This judgmental bias can prevent important information received later to be taken into account.
Ascertainment	Tendency to selectively analyse clinical data in the light of prior expectations or beliefs (belief bias). This bias can impact the interpretation of new information resulting from precise surveillance or screening of certain symptoms.
Availability bias	Tendency to form an opinion based on the most recent and readily available information in one's mind, considered more likely. For example, for an opinion on a treatment, we remember the last few patients rather than a series of 100.
Base-rate neglect	Type of error due to poor knowledge of disease incidence rates, either by underestimating or by overestimating the occurrence of a diagnosis.
Confirmation bias	Tendency to select and interpret information confirming a clinical intuition or a priori diagnosis, and to neglect information that contradicts or invalidates this intuition.
Diagnosis momentum	Diagnosis or treatment plans established by previous clinicians are rarely questioned by new practitioners and stick to the patient. This phenomenon can prevent considering new options and enhancing the diagnosis or provided healthcare.
Illusory correlation	Tendency to infer causation relationships between correlated but independent events.
Premature closure	Tendency to stop reasoning, evaluating or looking for a better diagnosis or treatment alternative after finding a suitable enough option (close to 'satisfaction search bias').
Primacy effect	Mnemonic bias, tendency to remember and consider more the first information out of a list of equal importance.
Recency effect	Mnemonic bias, tendency to remember and consider more the most recent information (received last), for example the last words of a clinical interview or the last symptoms of a list.
Unpacking principle bias	Type of error occurring when not all the necessary information were requested to make an objective judgement. The risk would be, for example, to omit information that would allow a differential diagnosis.
Affect bias	When decisions are made in a context where the immediate emotions are strong and can influence our choices.
Ambiguity or risk aversion	Type of bias describing the tendency to favour choices with known risks and associated probabilities rather than ambiguous or uncertain options.
Commission bias	Tendency to favour action over inaction, even when inaction would be more rational. It can result in overprescription.
Default bias or status quo bias	Tendency to stick to the default option and avoid changes. The cost of change in terms of cognitive effort is automatically considered too great and one continues to behave in the same way.
Framing bias	The perception of a situation can be influenced by the way options are being presented (formulation with different numerical presentations, or with positive or negative connotations ...).
Information bias	This bias translates into errors in the collection of information, for example during an interview: it can be a failure to observe, a misclassification or organisation of data, or errors in memory recall during synthesis.
Loss aversion	Tendency to be more sensitive to the loss of a certain amount of resources (cognitive effort, time, money ...) than to the gain of the same amount of resources, resulting in choices that tend to avoid losses rather than attempt gain.
Omission bias	Tendency to favour inaction or to avoid difficult issues over action ('wait and see'). It affects self-doubting clinicians.
Outcome bias	Tendency to focus on the outcome of the decision rather than the information to be interpreted to make a relevant decision. This bias is more common among clinicians with lower self-confidence and can lead to an incorrect diagnosis.

Representativeness restraint bias

Tendency to rely on the 'frequency argument,' i.e., to favour the most common hypotheses and not to mention the rarer ones. It is a restriction of thought that prevents a broader questioning of a clinical situation.

Retrospective prejudice

When the result of a situation is known, it can influence the way in which we perceive the preceding events as we forget the uncertainty we were facing at that time, and lead to fallacious reconstruction ('we are remaking history'). It can prevent learning and lead to the repetition of error.

Self-served bias

Tendency to reduce the analysis of clinical data and the diagnosis to one's own point of view. It affects communication between the different parties (physician, patients, and other stakeholders).

Sunk cost fallacy

Tendency, when one has already invested a lot of resources (time, energy or money) in a project or an action that seems to have little chance of succeeding, to continue investing although it is doomed to failure. In medicine, it is a question of pursuing an ineffective strategy, for example. Tendency to conform and reproduce a behaviour or an attitude just to act as others do.

Bandwagon effect

Tendency to conform and reproduce a behaviour or an attitude just to act as others do.

Fundamental attribution error

While making judgments about people's behaviour, it's the tendency to overemphasise dispositional factors or personality-based explanations and underestimate situational ones. The consequence is the risk of making incorrect judgments, discounting reasons that might have contributed to their observed behaviour.

Stereotyping

Tendency to infer characteristics about an individual based on the group in which we categorised him/her. This can result in a wrong diagnosis solely based on our belief that the patient belongs to a certain group with a typical disease.

[Mouchabac et al. 2021](#)

NUMERIC SYMPTOMS

“subjects have no need to be equipped with multiple sensors or even be burdened by **invasive devices** (e.g., endoscopy) [...] Additionally, CA can make it feasible to collect data from subjects via mobile devices (e.g., a smartphone), which can provide the subjects **24×7 monitoring service.**”

Qian et al. 2020

- + **no judgment**
- + **easy to access** (smartphones)
- + **no biases (patients and clinicians)**



Yes.

NUMERIC SYMPTOMS

“subjects have no need to be equipped with multiple sensors or even be burdened by **invasive devices** (e.g., endoscopy) [...] Additionally, CA can make it feasible to collect data from subjects via mobile devices (e.g., a smartphone), which can provide the subjects **24×7 monitoring service.**”

Qian et al. 2020

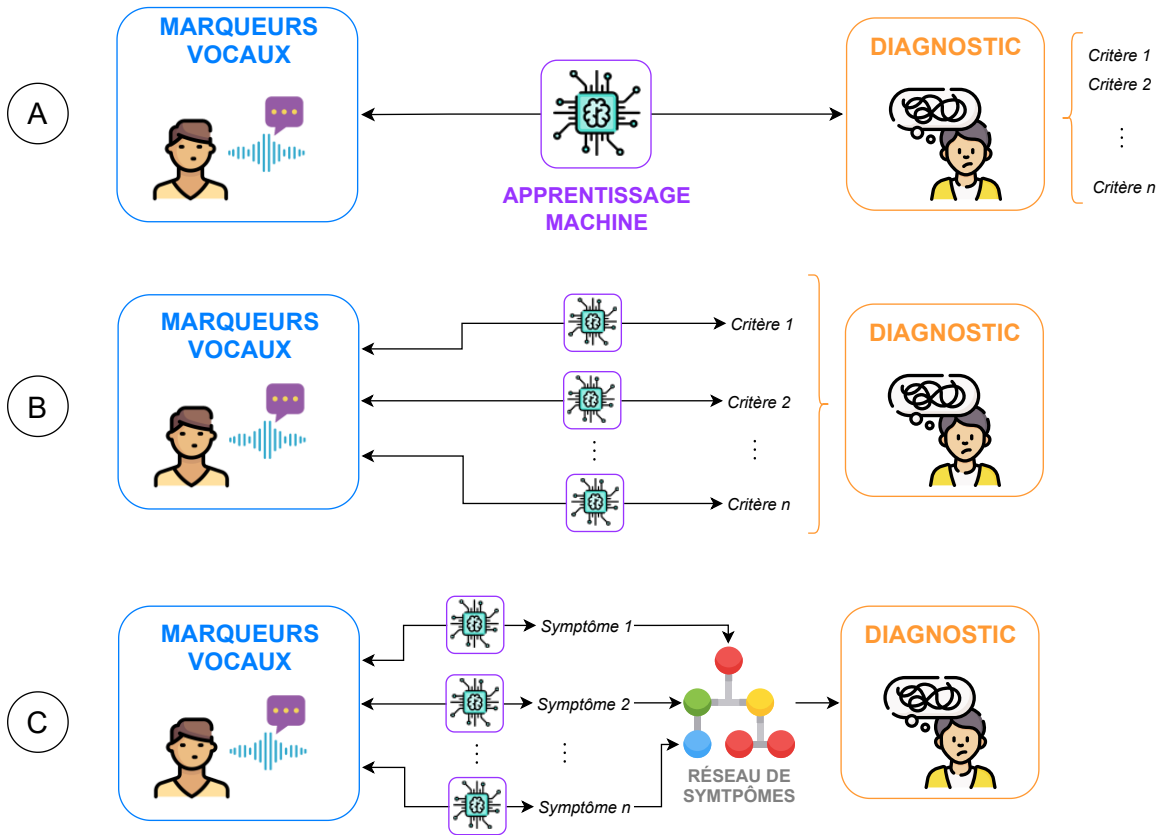
- + **no judgment**
- + **easy to access** (smartphones)
- + **no biases (patients and clinicians)**
- + **epistemic injustices**



Yes.

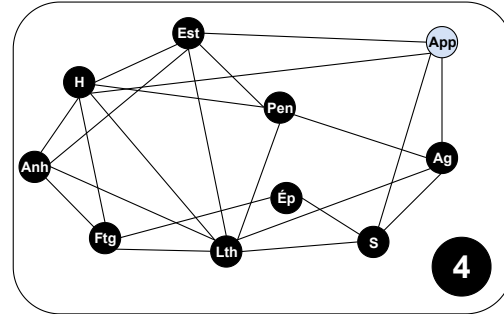
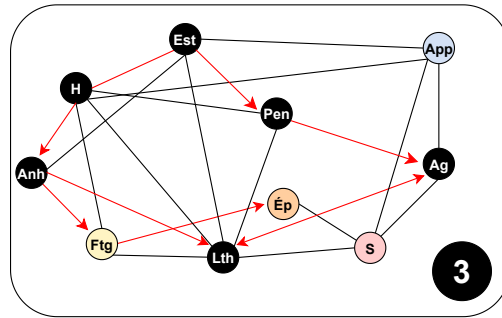
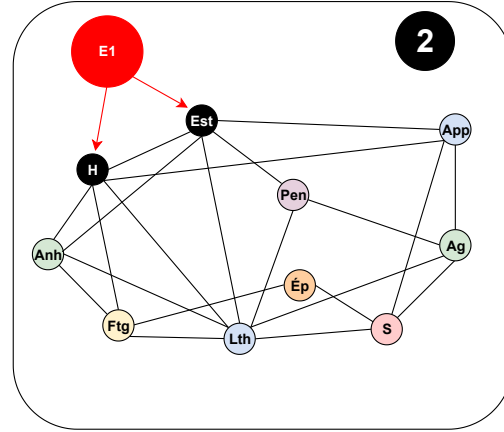
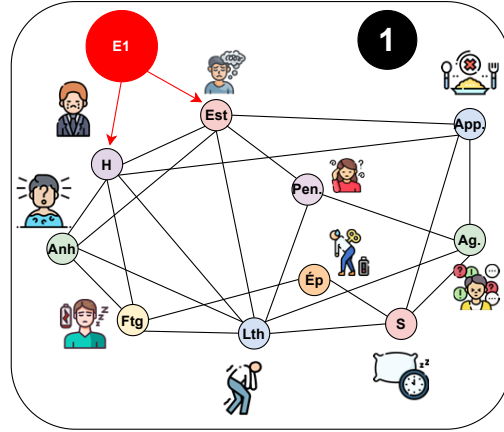
SYMPTOMS NETWORKS

48



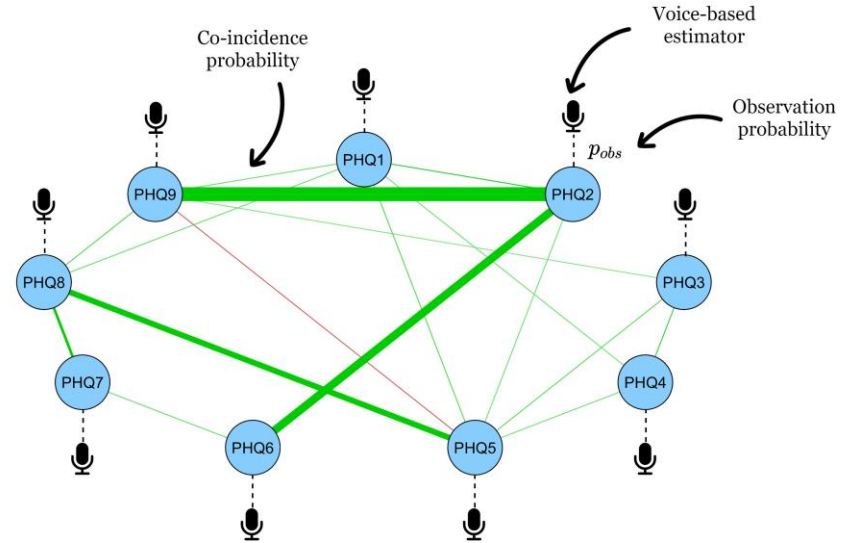
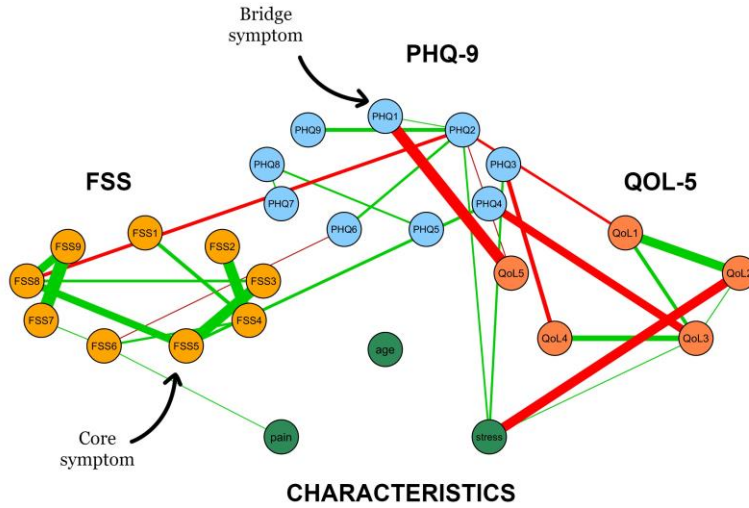
SYMPTOMS NETWORKS

49



SYMPTOMS NETWORKS

50



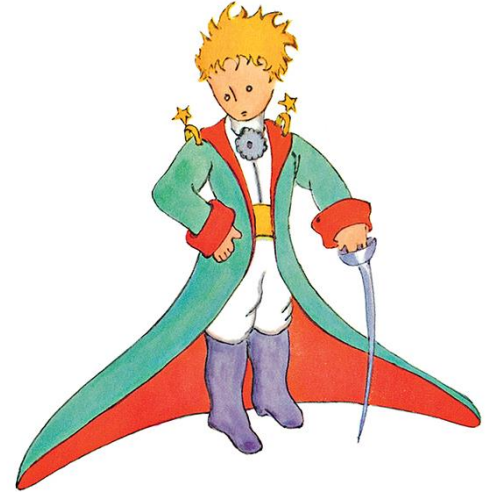
Conclusion

Doggy bag

DOGGY BAG

- ▶ Beware of **preconceived ideas** about medical practice when you are a computer scientist (get trained!)
- ▶ **Objectivity** is not necessarily better than subjectivity
- ▶ **Symptoms** instead of diagnoses

Thank you for
your attention!



QUESTIONS?



vincent.martin@labri.fr



[@V_P_Martin](https://twitter.com/V_P_Martin)



[Vincent-P-Martin](https://www.researchgate.net/profile/Vincent-P-Martin)

BONUS : « BUT IT WORKS ! »

54

▸ Adversarial attacks



"Panda"

57.7% de confiance

+ .007 ×



Bruit

BONUS : « BUT IT WORKS ! »

55

- ▶ Adversarial attacks
- ▶ Digression: music information retrieval
 - ▶ 2013 MIREX challenge « Audio Latin Music Genre classification task »

[Sturm 2016](#)



BONUS : « BUT IT WORKS ! »

56

	ChaCha	Jive	Quickstep	Rumba	Samba	Tango	Waltz	Pr
ChaCha	87.88	0.00	0.00	0.00	0.00	4.00	0.00	96.67
Jive	0.00	88.89	0.00	3.45	0.00	0.00	0.00	94.12
Quickstep	6.06	0.00	95.83	3.45	0.00	0.00	0.00	88.46
Rumba	6.06	0.00	4.17	75.86	0.00	0.00	0.00	88.00
Samba	0.00	5.56	0.00	3.45	96.00	0.00	0.00	92.31
Tango	0.00	5.56	0.00	0.00	0.00	96.00	0.00	96.00
Waltz	0.00	0.00	0.00	13.79	4.00	0.00	100.00	91.23
F	92.06	91.43	92.00	81.48	94.12	96.00	95.41	91.49

Time stretching
32 values
in [0.85, 1.15]



	ChaCha	Jive	Quickstep	Rumba	Samba	Tango	Waltz	Pr
ChaCha	41.11	1.35	21.28	16.50	11.68	19.03	1.40	42.26
Jive	1.15	33.63	7.55	8.67	12.00	2.27	1.20	40.51
Quickstep	10.40	9.31	25.45	12.49	2.70	1.08	10.03	29.09
Rumba	13.60	22.52	31.64	32.62	16.54	9.62	22.66	21.54
Samba	15.15	15.02	0.34	2.89	38.27	13.30	2.18	42.24
Tango	12.12	6.46	2.82	5.31	11.68	37.62	0.94	46.59
Waltz	6.47	11.71	10.92	21.53	7.14	17.08	61.59	62.57
F	41.68	36.75	27.15	25.95	40.16	41.63	62.07	38.61

BONUS : « BUT IT WORKS ! »

57

- ▶ Adversarial attacks
- ▶ Digression: music information retrieval
- ▶ The horse of Dr. Von Osten (Pfungst, 1911)



BONUS : « BUT IT WORKS ! »

58

- ▶ Adversarial attacks
- ▶ Digression: music information retrieval
- ▶ The horse of Dr. Von Osten (Pfungst,1911)

Solution ?
Irrelevant transformations

Sturm 2014



THE END

1. IS19 challenge: winner = Fischers vectors + SVR
Recent DL models : perf < IS19
 2. C. Rudin 2019 « Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead », *Nature Machine Intelligence*
 3. Meta analyses : Christodoulou « A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models » *J. Clinic. Epidemio.*
- Did you put as much efforts in logistic regression than in tuning a deep learning model?

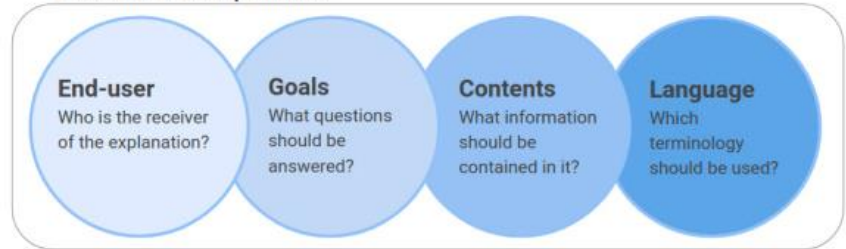
IS EXPLAINABILITY ENOUGH?

Vilone et al. 2021 “Notions of explainability and evaluation approaches for explainable artificial intelligence”, Information Fusion

Algorithmic transparency	The degree of confidence of a learning algorithm to behave ‘sensibly’ in general [2], [26]
Causality	The capacity of a method for explainability to clarify the relationship between input and output [8], [21], [22], [23], [24], [25], [29]
Comprehensibility	The quality of the language used by a method for explainability [9], [31], [32], [33], [34], [35], [36], [37], [38]
Effectiveness	The capacity of a method for explainability to support good user decision-making [40], [41], [42], [43]
Efficiency	The capacity of a method for explainability to support faster user decision-making [20], [41], [42]
Explicability	The degree of association between the expected behaviour of a robot to achieve assigned tasks or goals and its actual observed actions [44]
Explicitness	The capacity of a method to provide immediate and understandable explanations [45]
Faithfulness	The capacity of a method for explainability to select truly relevant features [45]

Interestingness	The capacity of a method for explainability to facilitate the discovery of novel knowledge and to engage user’s attention [33], [34], [36], [53], [54]
Interpretability	The capacity to provide or bring out the meaning of an abstract concept [9], [18], [33], [35], [55], [56], [57], [58], [59], [60], [61], [62], [63]
Informativeness	The capacity of a method for explainability to provide useful information to end-users [21]
Justifiability	The capacity of an expert to assess if a model is in line with the domain knowledge [1], [33], [40], [55], [64], [65]
Mental Fit	The ability for a human to grasp and evaluate a model [33], [66]
Persuasiveness	The capacity of a method for explainability to convince users perform certain actions [20], [41], [42]
Selection/ simplicity	The ability of a method for explainability to select only the causes that are necessary and sufficient to explain the prediction of an underlying model [25]
Soundness	The extent to which each component of an explanation’s content is truthful in describing an underlying system [27], [28]
Transparency	The capacity of a method to explain how the system works even when it behaves unexpectedly [9], [10], [11], [12], [20], [26], [40], [41], [47], [58], [59], [63], [64], [76], [77], [78]

Structure of an explanation

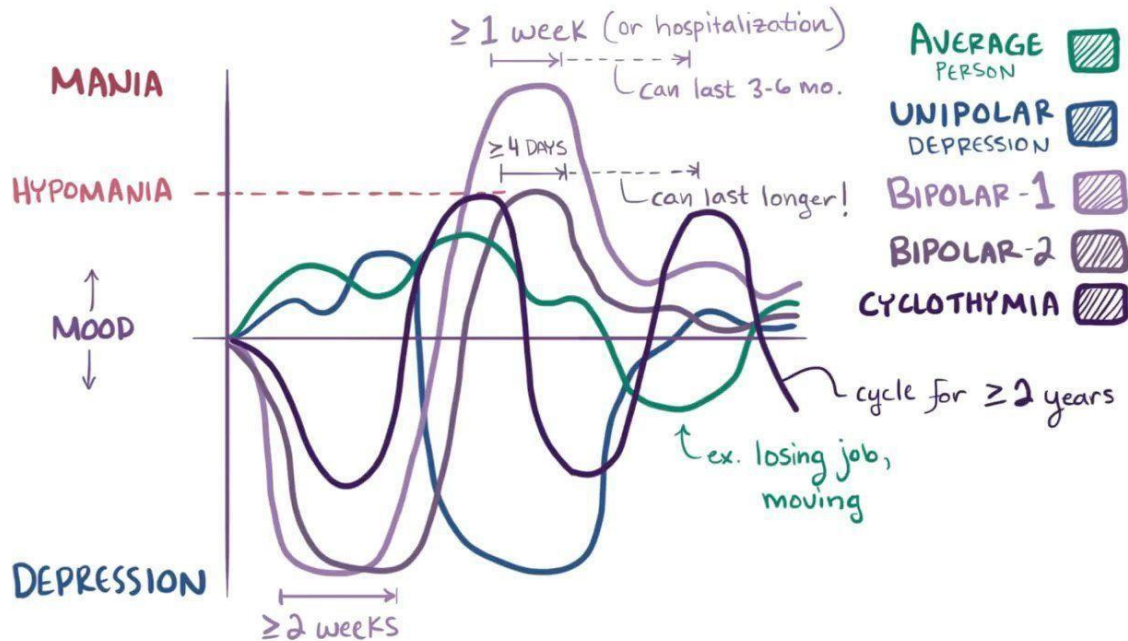


WHAT DO WE DETECT ?

EXAMPLE 1: BIPOLAR DISORDER

Bipolar disorders

- ▶ Diag. = based on **variations** and **duration**
- ▶ How to detect BD with only **1 recording ?**
- ▶ **State** vs. **Trait**
(same for all Ψ disorders)



https://www.osmosis.org/learn/Bipolar_disorder

WHAT DO WE DETECT?

EXAMPLE2: DEPRESSION

What does a ML classifier learn ?

- ▶ **Difference between groups**
 - ▶ Sub-group?
 - ▶ Symptom?
 - ▶ Other bias?
 - ▶ ?
- ▶ + **/!\ Temporalty /!**

Depressive vs. **HC with bad mood**

- ▶ **NOT “depression disorder”**

