INVENTEURS DU MONDE NUMÉRIQUE

Multimodal Data in the Medical Domain

Dr Alexandra König, neuropsychologist
Email: alexandra.konig@inria.fr

Background…

- Clinical insights through keen behavioral observation & subjective questionnaires and scales
- Punctual assessments based on interactions only
- Variability in results, errors and misdiagnosis

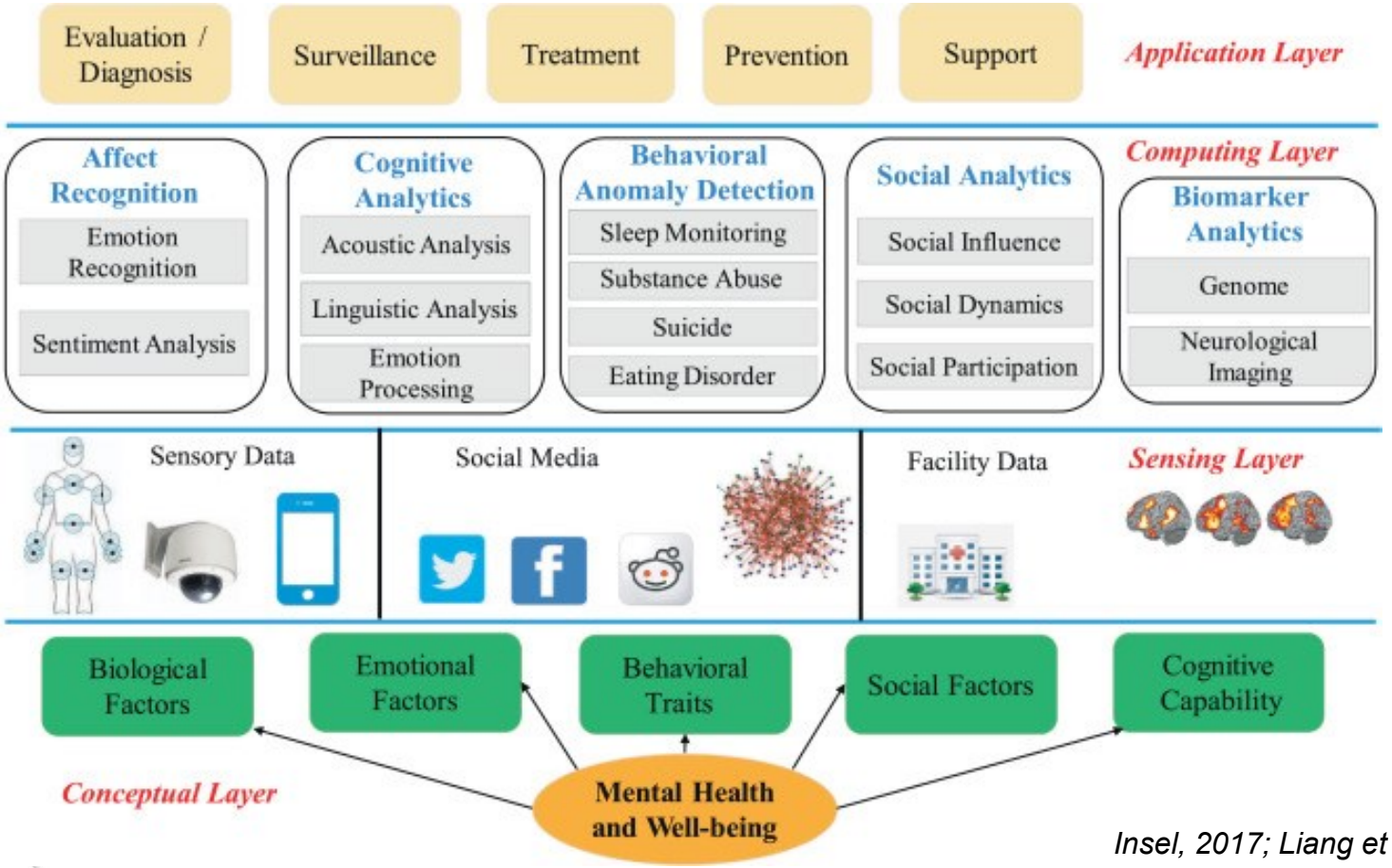
→ Need for diverse data and dimensional approaches as fundamental principles for better diagnosis of mental disorders

→ “measurement-based care” rather than relying only on clinical judgment

→ Shift from late disease management and cure to preventative-personalized precision medicine

Digital phenotyping

"moment-by-moment quantification of the individual- level human phenotype in situ using data from personal digital devices , especially smartphones and wearable sensors "



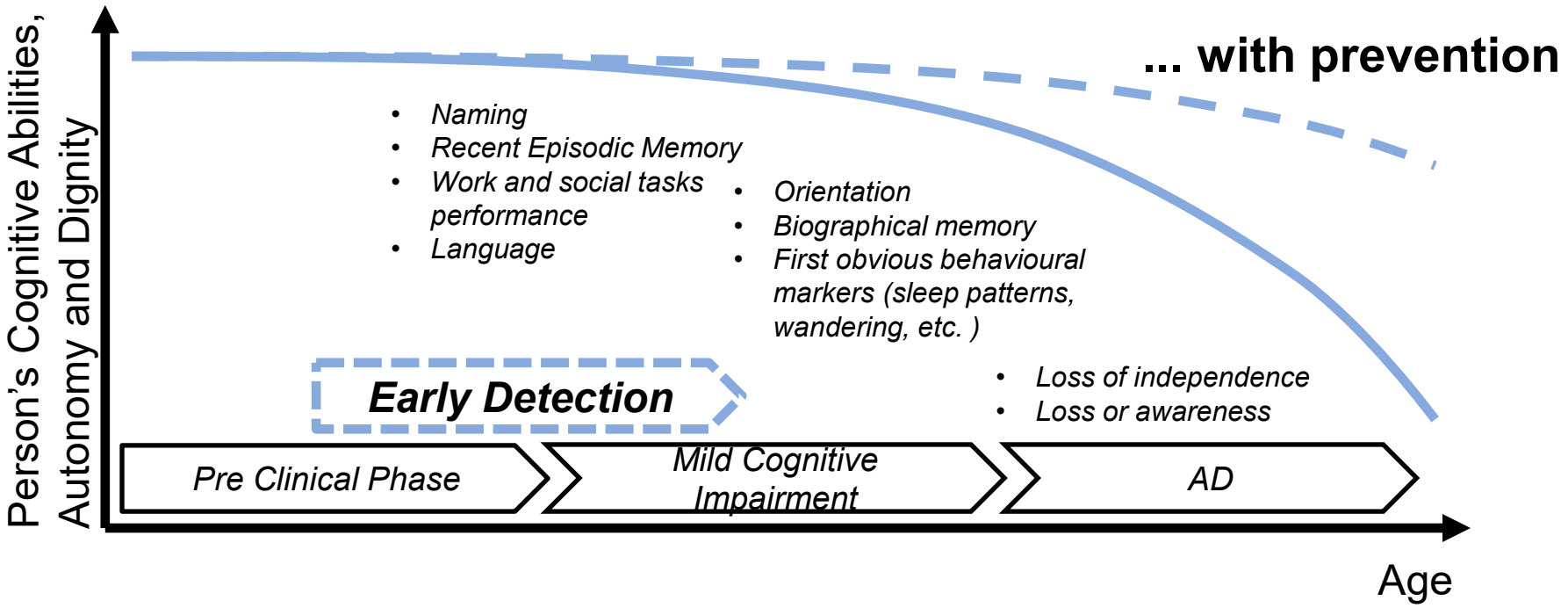
Insel, 2017; Liang et al., 2019.

Research project examples

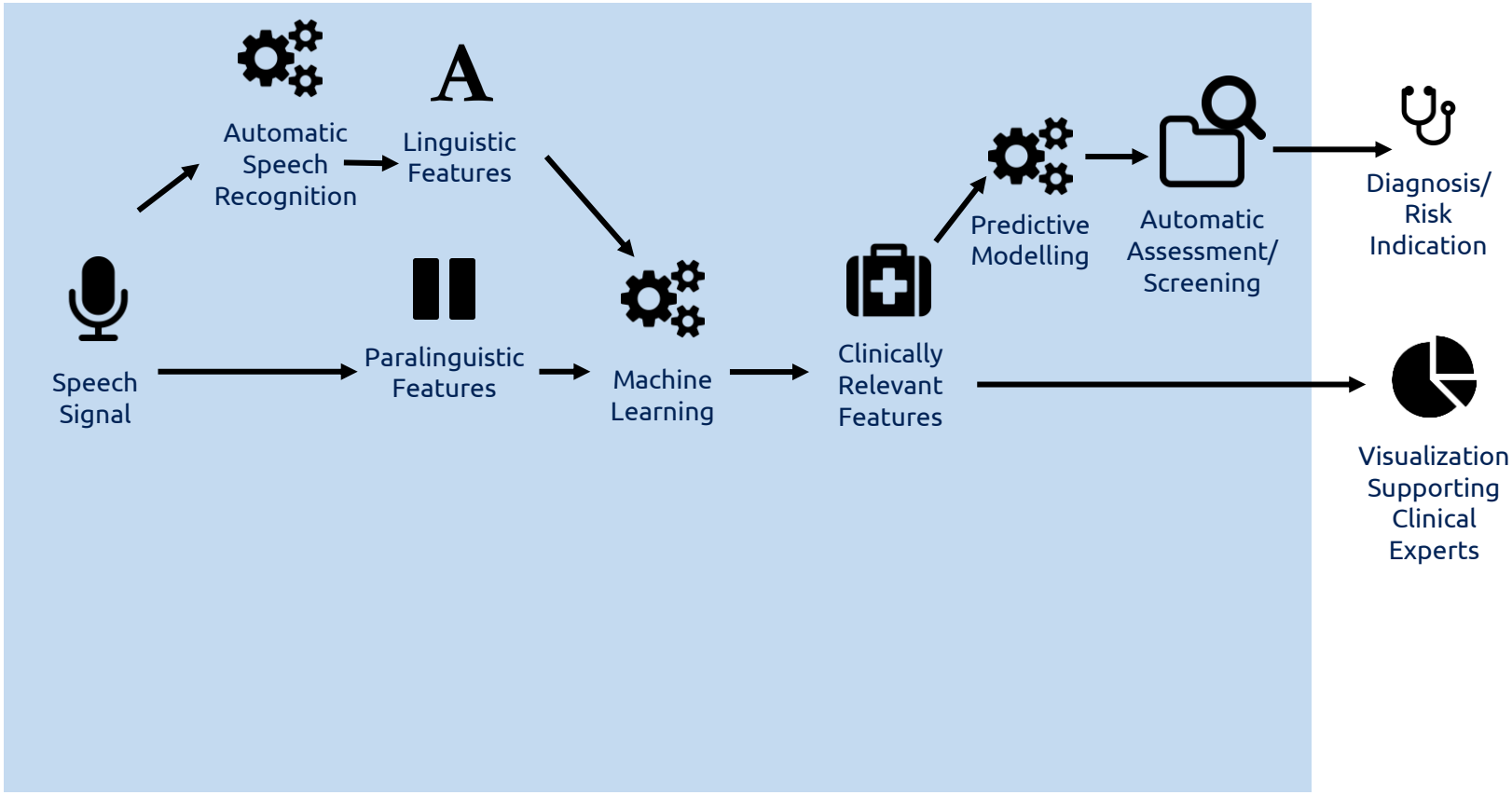
- Deep Speech Analysis for Cognitive Assessment in Clinical Trials (**DeepSpa**) / European Insitute for Innovation & Technology (EIT) Health

- **Mephesto** - Digital Phenotyping 4 Psychiatric Disorders from Social Interaction/ Joint Program INRIA & DFKI

Neurodegenerative Diseases (Alzheimer's)



Speech Analysis platform



Use case 1: Telephone-based pre-screening

Need for:
- automatic
- inexpensive
- remote

screening solutions

Recorded speech is analysed by:
- Linguistic
- Paralinguistic features

Battery of speech-based tasks
Subjects:
- Subjective CI
- Mild CI
- Dementia

Machine classification of clinically relevant features

Fully automated analysis of phone-based assessment is feasible for dementia screening

Classification results reveal good sensitivity

The word error in automated speech recognition is acceptable

DeepSpA Study: Longitudinal remote-monitoring of SCI

Maastricht University

120 participants, Age ~ 60, Subjective Cognitive Decline (SCI)

Baseline F2F assessment 1
Δelta clinical

- Audio Recordings Cognitive Tests
  - Verbal Learning Test (V1)
  - Verbal Fluency (V1)

+ biomarkers (gold standard)
+ consents

Semi-automated telephone call

+ verbal learning (V2)
+ digit span
+ verbal fluency
+ questionnaire SCI

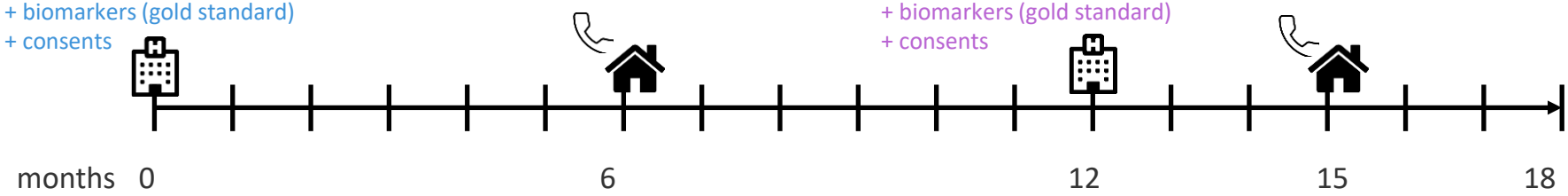
Baseline F2F assessment 2
Δelta clinical

- Audio Recordings Cognitive Tests
  - Verbal Learning Test (V3)
  - Verbal Fluency (V2)

+ biomarkers (gold standard)
+ consents

Fully-automated telephone call

+ verbal learning
+ verbal fluency

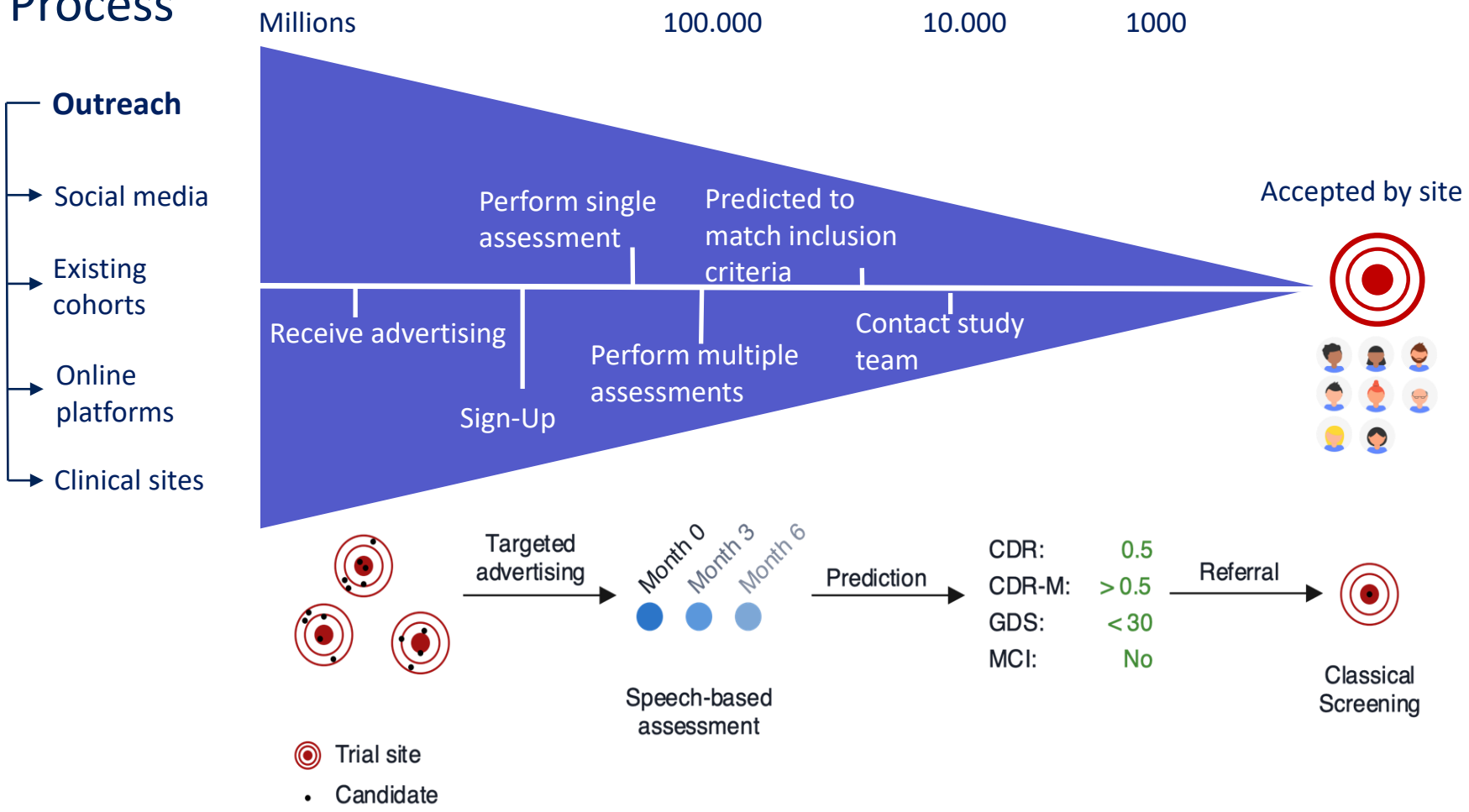


DeepSpA: Face2Face & Phone scores correlation

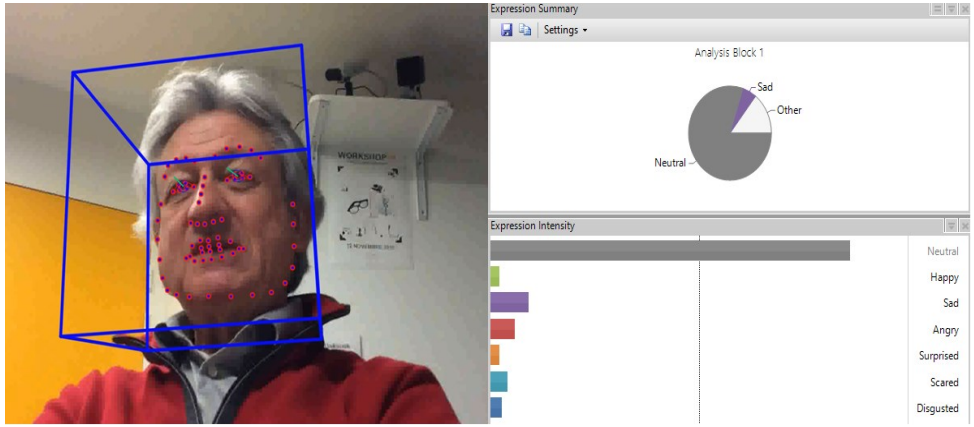
		Clinical				Baseline			M6	
		CDR SOB	CDR Tot	MMSE	IADL	Word count	Word count	Norm score	Word count	Norm score
Clinical	CDR SOB	1								
	CDR Tot	.74	1							
	MMSE	.49	.40	1						
	IADL	.72	.49	.39	1					
Baseline	Word count	.33	.30	.40	.39	1				
	Word count	.41	.32	.48	.44	.87	1			
	Norm score	.38	.29	.36	.37	.74	.86	1		
M6	Word count	.40	.22	.40	.30	.73	.50	.36	1	
	Norm score	.37	.33	.34	.19	.56	.23	.27	.88	1

-> high agreement between Face2Face and phone assessment

Process


Millions
100.000
10.000
1000
Outreach
Social media
Existing cohorts
Online platforms
Clinical sites
Receive advertising
Sign-Up
Perform single assessment
Perform multiple assessments
Predicted to match inclusion criteria
Contact study team
Accepted by site
Targeted advertising
Month 0
Month 3
Month 6
Speech-based assessment
Prediction
CDR: 0.5
CDR-M: > 0.5
GDS: < 30
MCI: No
Referral
Classical Screening
Trial site
Candidate

Combining audio & video for behavioral analysis



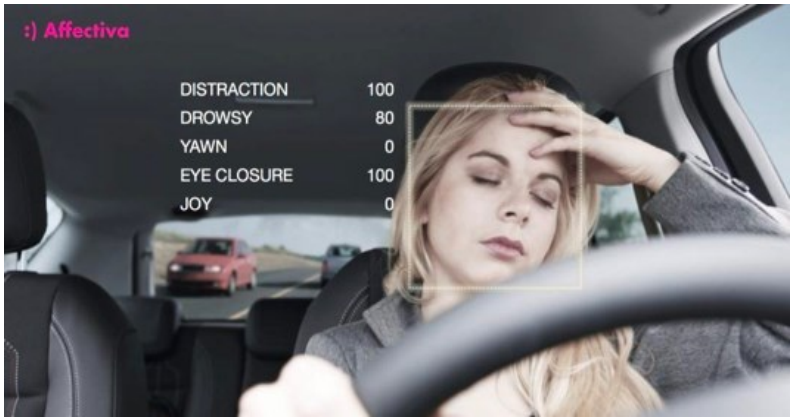
Automatic analysis of facial expression to capture :

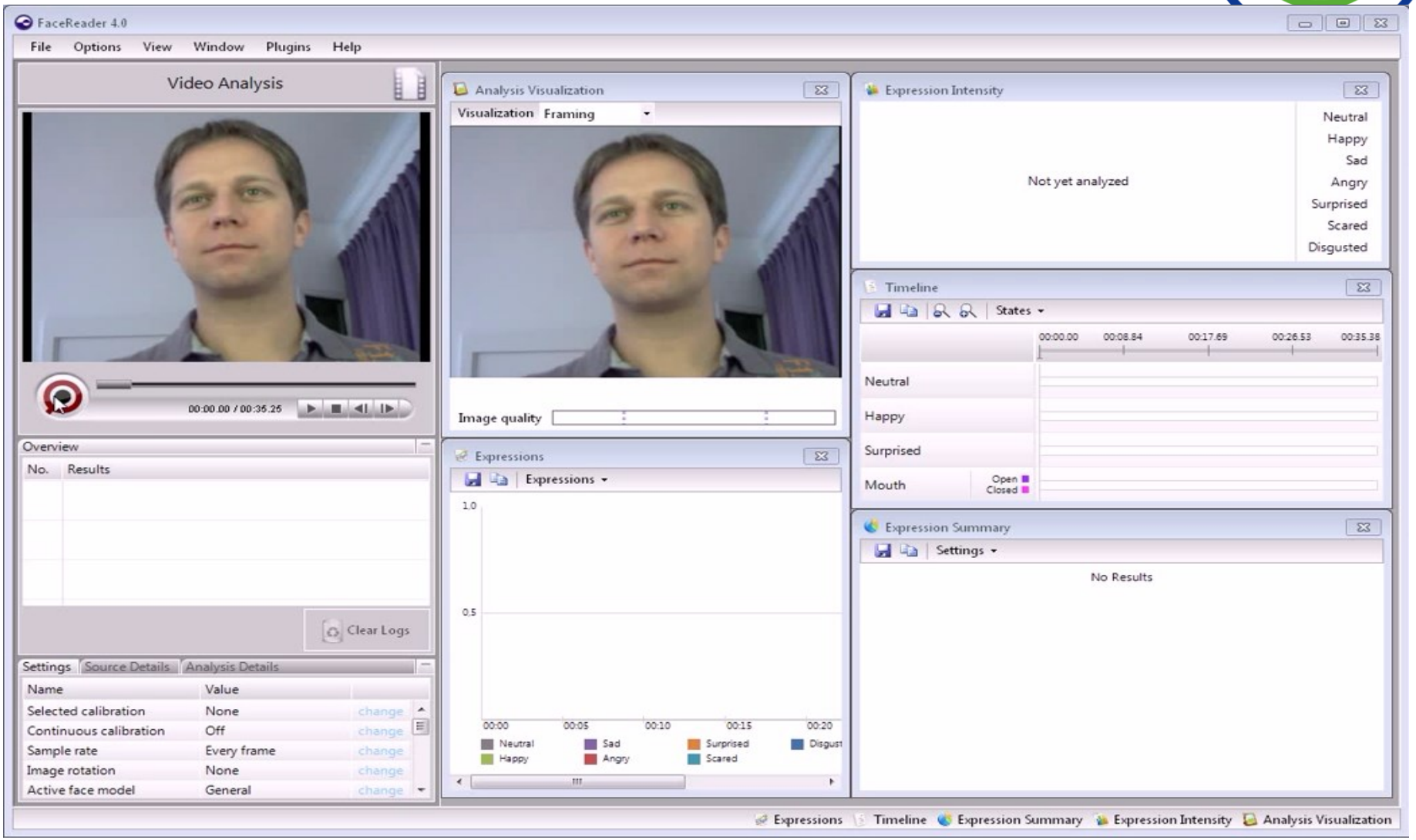
- Emotional states/ Psychological symptoms
- Levels of engagement/levels of focus on task
- Weariness/Tiredness
- Control for cheating

- Specific face action unit abnormalities, mutual gaze and vowel space in speech

→ highly associated with depression severity *(Baker et al., 2016)*

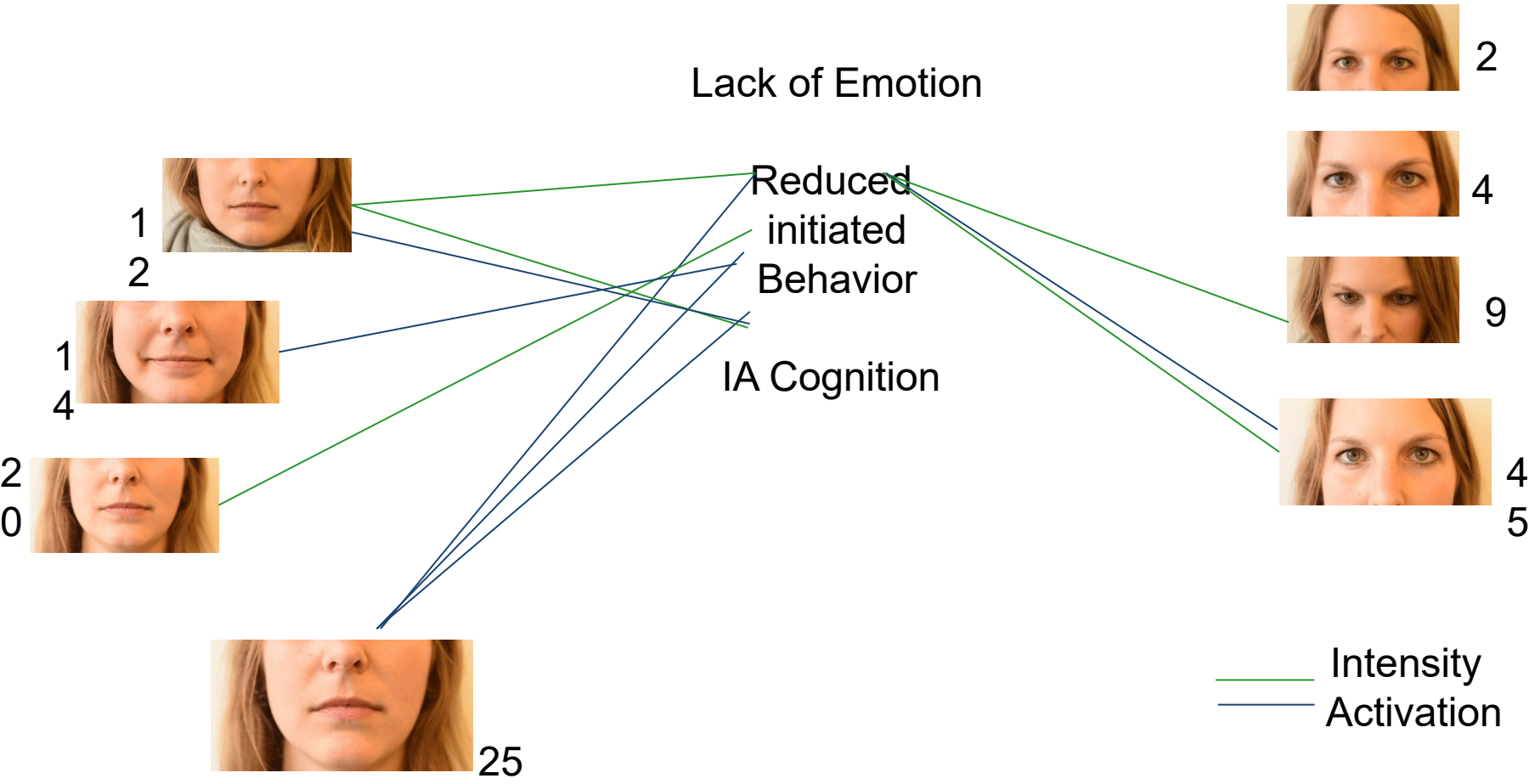
- Multimodal fusion of acoustic and visual measures better classification results (close to 90%) in accuracy than unimodal classification performance *(Scherer, 2013)*




FaceReader 4.0
File
Options
View
Window
Plugins
Help
Video Analysis
00:00.00 / 00:35.25
Overview
No.
Results
Clear Logs
Settings
Source Details
Analysis Details
Name
Value
Selected calibration
None
change
Continuous calibration
Off
change
Sample rate
Every frame
change
Image rotation
None
change
Active face model
General
change
Analysis Visualization
Visualization
Framing
Image quality
Expressions
Expressions
1.0
0.5
00:00
00:05
00:10
00:15
00:20
Neutral
Happy
Sad
Angry
Surprised
Scared
Disgust
Expression Intensity
Not yet analyzed
Neutral
Happy
Sad
Angry
Surprised
Scared
Disgusted
Timeline
States
00:00.00
00:08.84
00:17.69
00:26.53
00:35.38
Neutral
Happy
Surprised
Mouth
Open
Closed
Expression Summary
Settings
No Results
Expressions
Timeline
Expression Summary
Expression Intensity
Analysis Visualization

Apathy assessment : Action unit mean during positive story telling

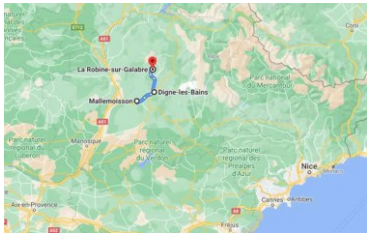
All negative correlations



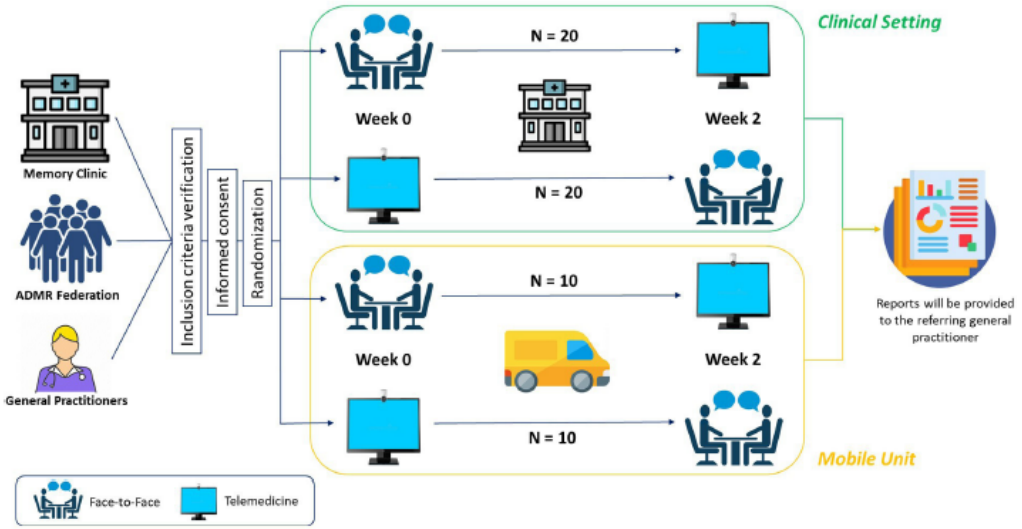
Use case II: Remote cognitive monitoring via telemedicine in rural areas

INRIA , City Digne-les-bains

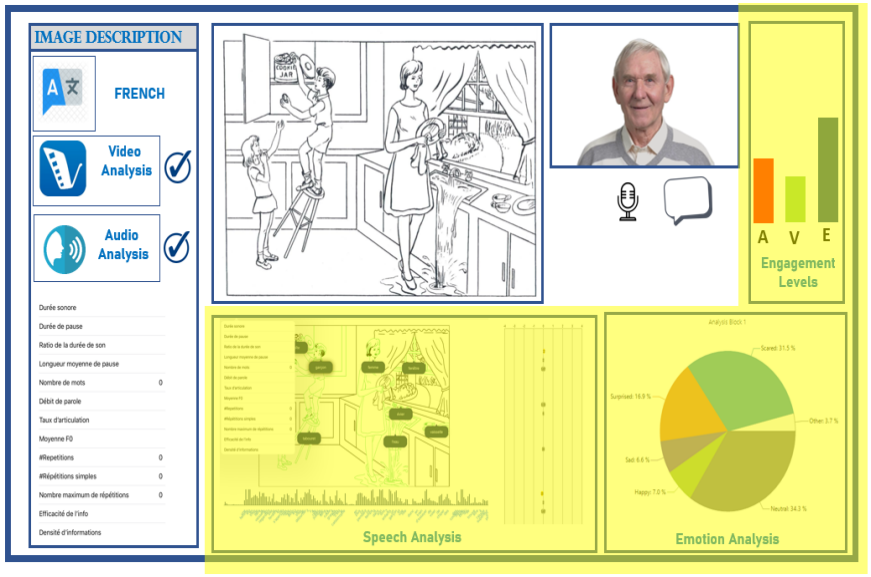
60 participants, Age ~ 55

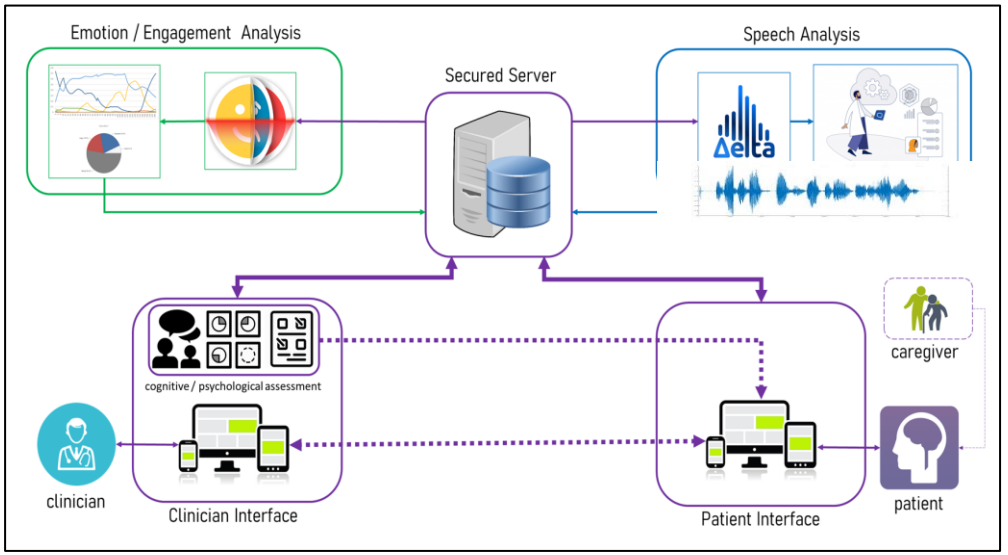


Study design



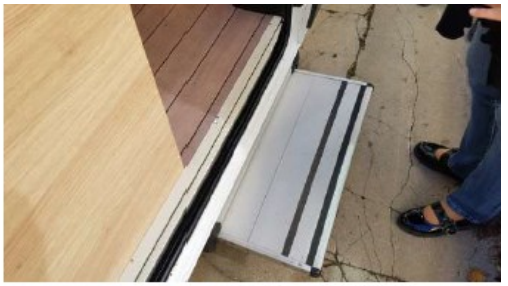
Interface

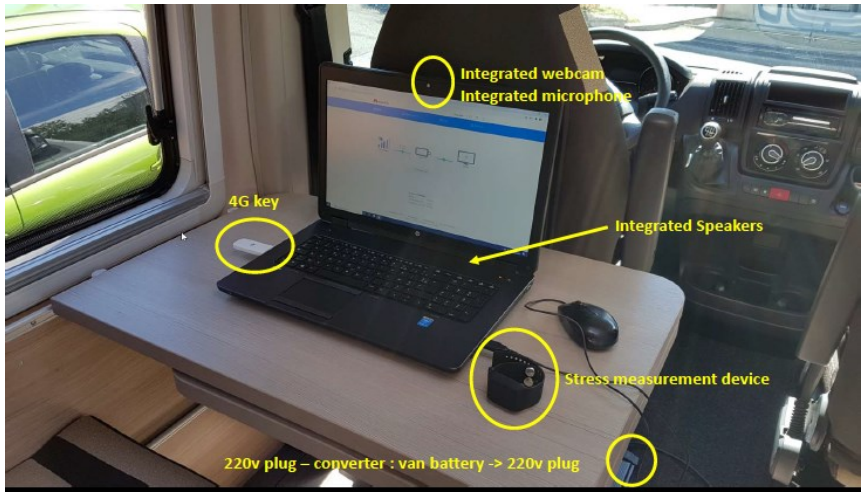


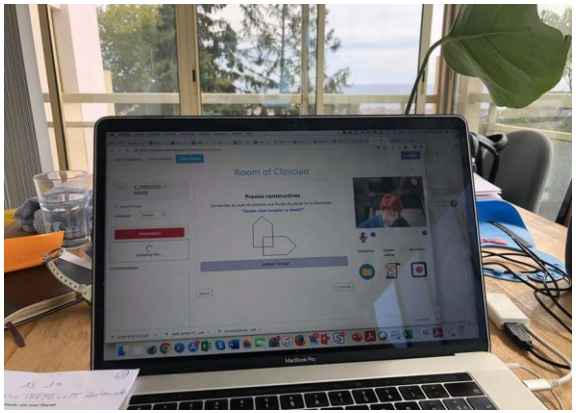


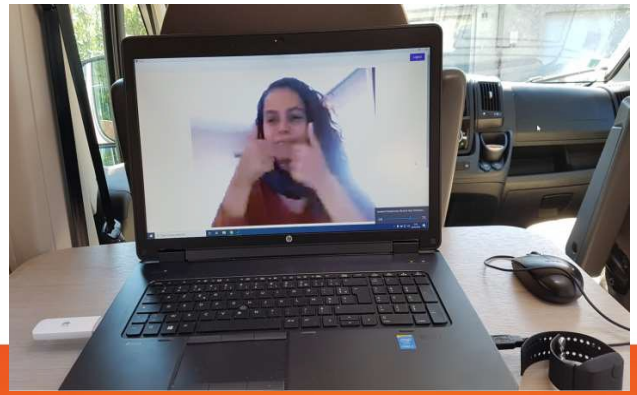
König et al., 2021

Mobile Unit



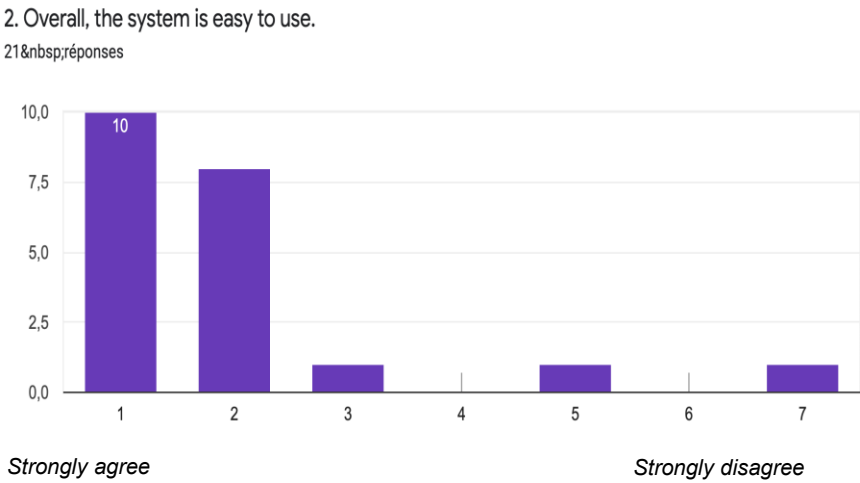

Integrated webcam
Integrated microphone
4G key
Integrated Speakers
Stress measurement device
220v plug – converter : van battery -> 220v plug


Room of Clinician

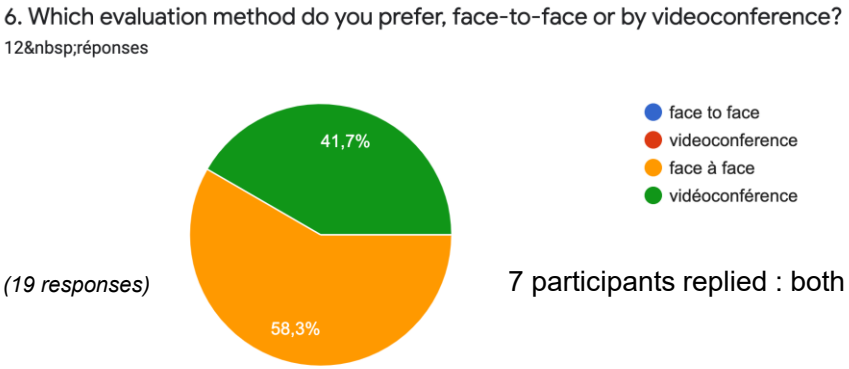


Results & User feedback

Cognitive tests	N	TNP	FTF
MMSE[1], total score	50	26,42	28,24
FCSRT			
Total recall score	42	43,91	42,91
Delayed recall	42	15,16	15,07
Recognition score	41	15,70	15,69
Lexis, total score	46	56,74	58,34
Stroop			
Reading, duration(s)	42	54,10	59,22
Interference, duration (s)	40	153,44	152,52
SVF (z-score)	47	-0,33	0,34



."Strangely, I feel more free in front of the computer to express myself"
"It still feels intimate and personal but this might depend on the clinician"
"For me it is not different than face to face and I'm used to it."
"not rather pleasantly surprised, puts a distance that is rather facilitating"



(19 responses)

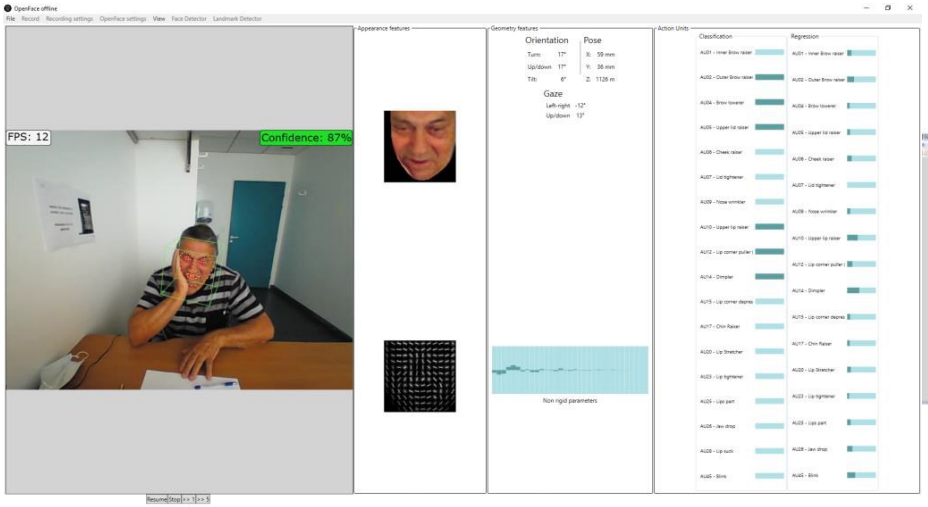
7 participants replied : both methods

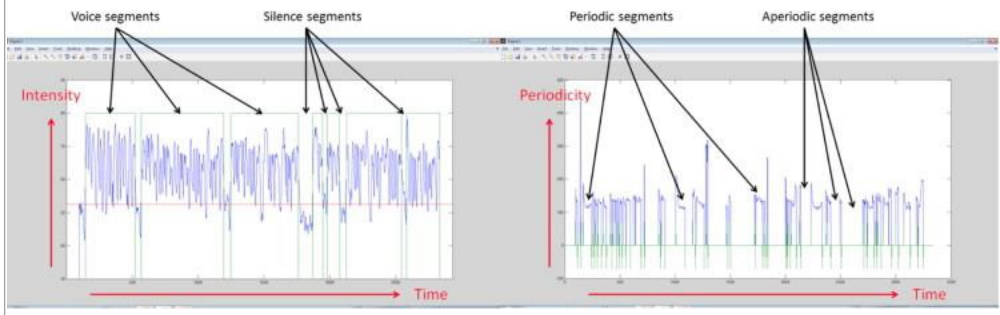
Patient 1: "Although I was a bit nervous for the phone-testing (always when something new happens), it was convenient that the testing could take place at home. The instructions were clear (both content-wise and audio), but the test test-leader sounded as very-far away. I missed face-to-face contact, because now I had to get the cup of coffee myself."

Future work

- Analysis of body behaviors (*facial expressions, head/gaze directions, gestures, postures,…*) to support assessments with ***visual biomarkers*** : emotions, engagement, fatigue, stress, level of concentration, focus ….

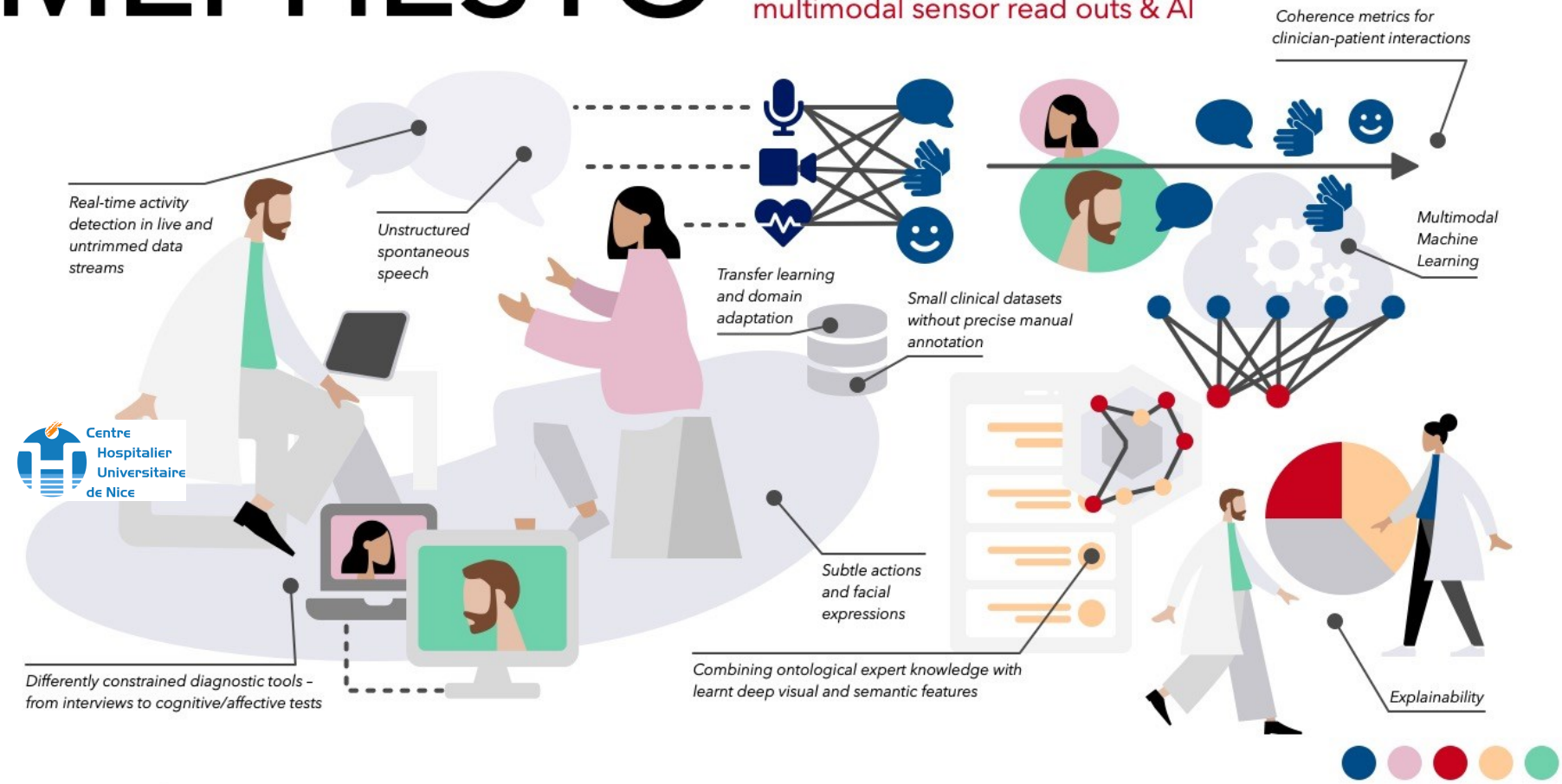
- Voice and ***speech*** analysis for new ***biomarkers*** to provide new insights on the patients' state and / or validate certain conclusions





MEPHESTO

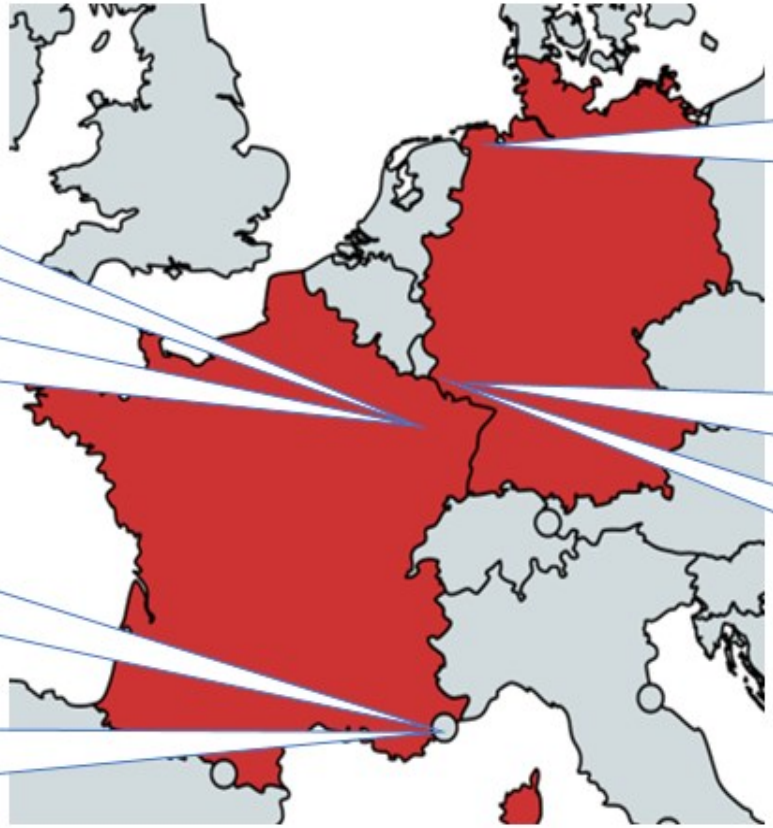
Digital interactional phenotyping |
Modelling clinical symptoms through multimodal sensor read outs & AI



Inria

German Research Center for Artificial Intelligence

Team



Nancy
SEMAGRAMME

Dialogue analysis

School of Medicine & Health Sciences
Carl von Ossietzky University of Oldenburg

UNIVERSITÉ DE LORRAINE

Psycholinguistic Studies

UKS Department of Psychiatry and Psychotherapy
Saarland University Hospital

Centre Hospitalier Universitaire de Nice

CBTEK Cognition Behaviour Technology

Hopital Pasteur, Psychiatry

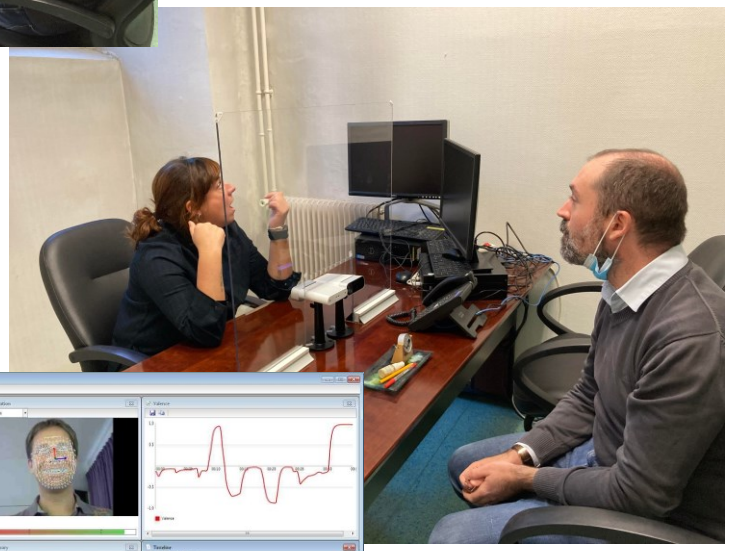
E.g. Team Competences

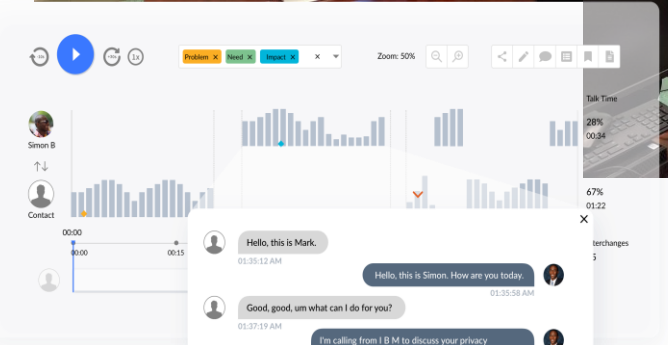
German Research Center for AI
Speech and language analysis

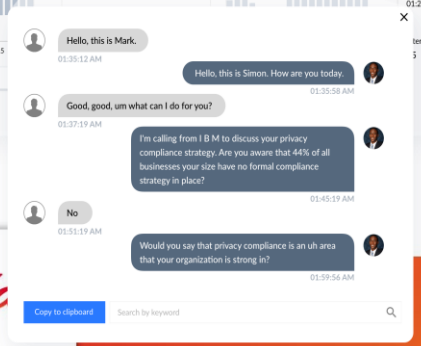
Sophia Antipolis
STARS
Video analysis

MEPHESTO

Technical Set up at Pasteur

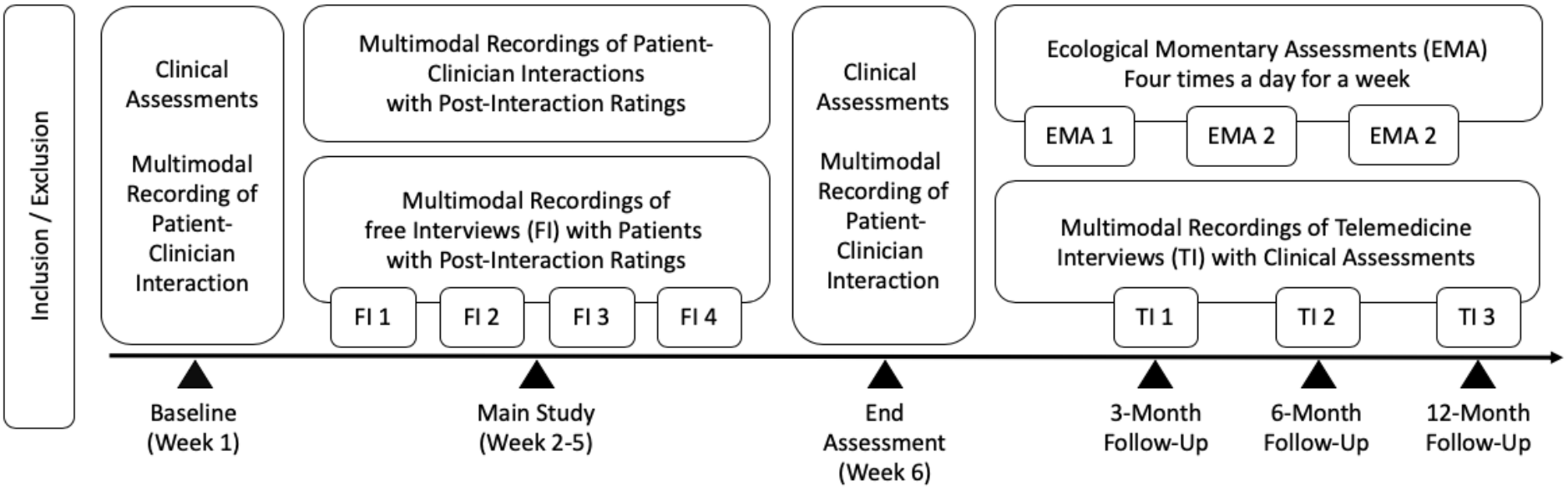






Mephesto study design

N= 150 (75 Major Depressive disorder / 75 Schizophrenia patients)



MEPHESTO

4 powerful clinical research cases that lead the way

→ Create a longitudinal multimodal corpus of patient-clinician interactions within the context of psychiatric disorders

Use case A: Supporting differential diagnosis for major depressive episode etiology

Use case B: Quantifying therapeutic alliance by means of social synchrony

Use case C: Treatment outcome/Relapse prediction from negative symptoms in schizophrenia

Use case D: Robust and objective measurement of formal thought disorder in schizophrenia

Thank you for your attention!

Questions?

alexandra.konig@inria.fr