



# REBECCA

Research on breast cancer induced chronic conditions  
supported by causal analysis of multi-source data

## Obtaining insights for the emotional well-being of cancer patients using NLP on their social media and browser history

Georgia Pantalona, Lazaros Apostolidis, Filareti Tsalakanidou,  
Symeon Papadopoulos



ITI - CERTH

March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# NLP and emotional well-being

## Diagnosis of mental issues



## Treatment assistance



Peng, Zhichao & Hu, Qinghua & Dang, Jianwu. (2019). Multi-kernel SVM based depression recognition using social media data. *International Journal of Machine Learning and Cybernetics*. 10. 10.1007/s13042-017-0697-1.

Mike Thelwall, TensiStrength: Stress and relaxation magnitude detection for social media texts, *Information Processing & Management*, Volume 53, Issue 1, 2017, Pages 106-121, ISSN 0306-4573,

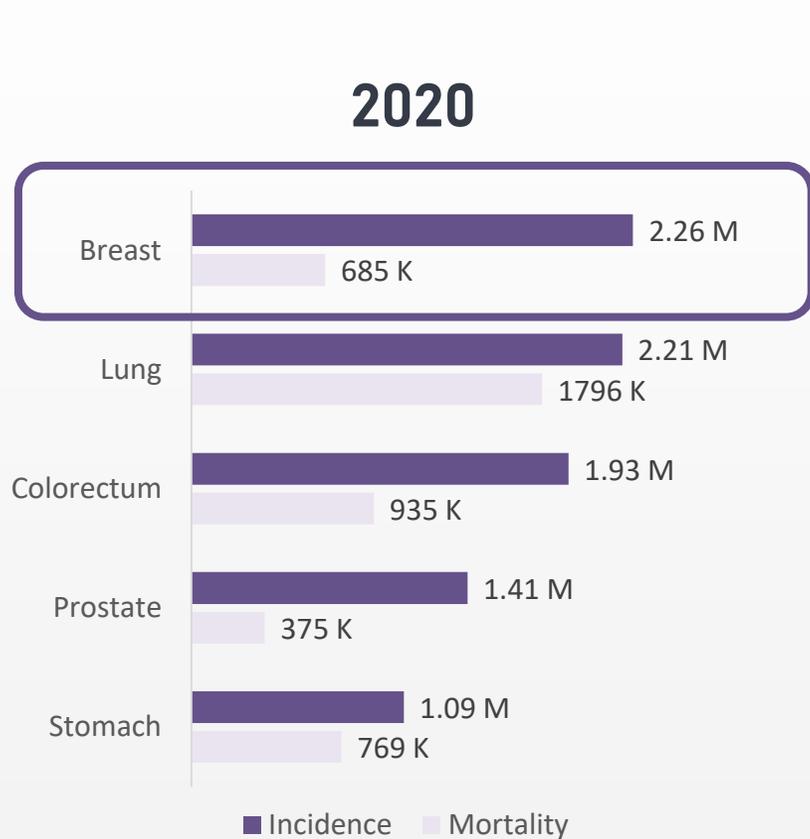


March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# Breast Cancer Worldwide



7.8B  
last 5 years

50%

Less than 64 years old

90%

Live for 5 years or more

Statistics from international Agency for Research on Cancer



March 28, 2022

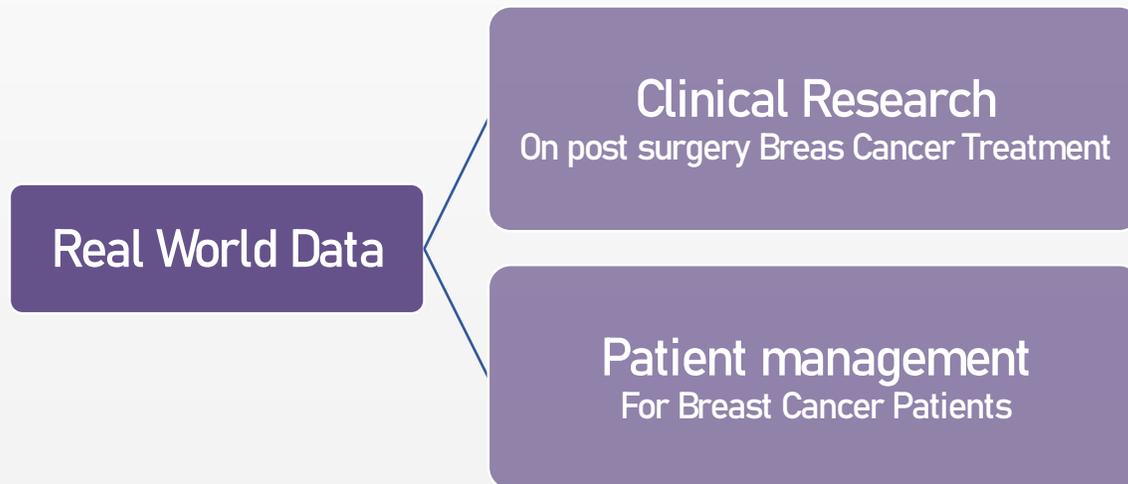


This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

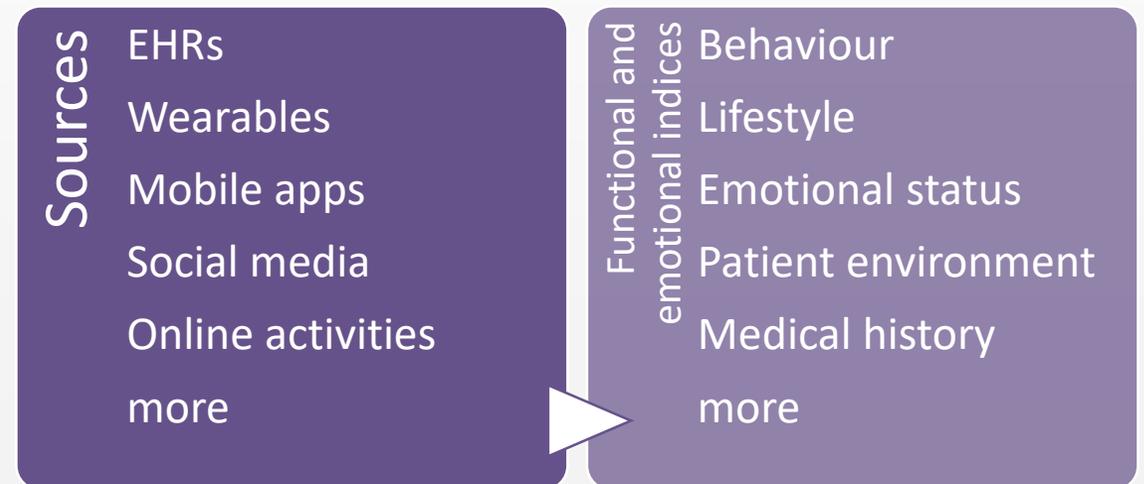
# REBECCA

Horizon 2020 project - REsearch on BrEast Cancer induced chronic conditions supported by Causal Analysis of multi-source data

## The Vision



## 360° patient monitoring using multi-source RWD



March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# Effects on emotional well-being

**GOAL**

Cancer can affect the patient's emotional well-being by causing:



Stress



Depression



Anxiety



Fear of recurrence

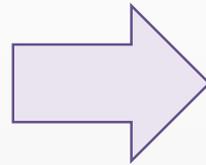
Generally affects their Quality of Life

Carreira H, Williams R, Dempsey H, Stanway S, Smeeth L, Bhaskaran K. (2021). Quality of life and mental health in breast cancer survivors compared with non-cancer controls: a study of patient-reported outcomes in the United Kingdom. *Journal of Cancer Survivorship*. 15. 1-12

Burgess C, Cornelius V, Love S, Graham J, Richards M, Ramirez A. (2005). Depression and anxiety in women with early breast cancer: five year observational cohort study. *BMJ*. 330: 702. *BMJ (Clinical research ed)*. 330. 702.

# Importance of emotional well-being

1. Good mental health is of great importance for the treatment trajectory <sup>[1]</sup>
2. Cancer recurrence is associated with poor mental health <sup>[2]</sup>



**Proper monitoring of the emotional state of the patients is crucial. It can assist experts to be aware and intervene early if necessary, by providing support or by other means <sup>[3]</sup>**

[1] M. Moreno-Smith, S. K. Lutgendorf, and A. K. Sood, "Impact of stress on cancer metastasis," *Future Oncol. Lond. Engl.*, vol. 6, no. 12, pp. 1863–1881, Dec. 2010, doi: 10.2217/fon.10.142.

[2] H. Götze, S. Taubenheim, A. Dietz, F. Lordick, and A. Mehnert-Theuerkauf, "Fear of cancer recurrence across the survivorship trajectory: Results from a survey of adult long-term cancer survivors," *Psychooncology.*, vol. 28, no. 10, pp. 2033–2041, 2019, doi: 10.1002/pon.5188.

[3] J. Singh Shrestha et al., "Social support, quality of life and mental health status in breast cancer patients," *Cancer Rep. Rev.*, vol. 1, no. 2, 2017, doi: 10.15761/CRR.1000107.

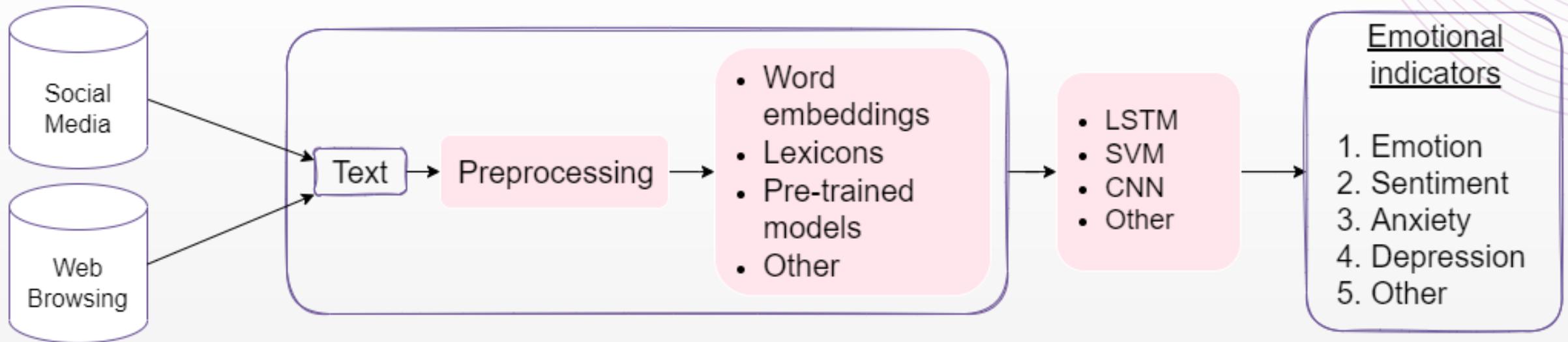


March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# Process - Pipeline



March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# Data Sources



Posts, comments, reading material

This data has the potential to **reveal** the user's **emotional state**.

For example, if the user is posting about events they have experienced either positive or negative



Use web browser histories

This data has the potential to reveal **mood, emotions, anxiety levels etc.**

For example, if someone constantly searches and reads about health-related topics then that's an indicator of anxiety and stress.



March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# Data gathering

## Google Chrome extension

- Collection of **pseudonymized** data from user's online activity
  - **Facebook** (Queries, comments, reactions, visits)
  - **YouTube** (Queries, comments, reactions)
  - **Web** (Queries, Websites visited)
- Control of monitoring (pause user tracking)
- Dashboard for deleting records user feels uncomfortable sharing

The screenshot shows the REBECCA dashboard interface. At the top, there is a header with the REBECCA logo and a 'Delete all my data' button. Below the header, there are two tabs: 'SETTINGS' and 'DATA'. The 'DATA' tab is active, showing a 'Filters' section with 'Set Dates' (Last Day, Last Week, Custom) and 'Set results' (10, 20, 50, 100). A search bar is present with a 'Search' button. Below the filters, there are two tables: 'Stored Urls' and 'Stored Queries'. The 'Stored Urls' table has columns for #, Url, Title, Session ID, Timestamp, and Tab. The 'Stored Queries' table has columns for #, Query, Session ID, Timestamp, and Tab.

#	Url	Title	Session ID	Timestamp	Tab
#1	https://rebeccaproject.eu.teamwork.com/#/messages/623637	REBECCA - Messages - REBECCA project	1638188948376_dt30lk	30 Nov 2021 17:22:18	existing
#2	https://rebeccaproject.eu.teamwork.com/#/messages/623637	REBECCA - Messages - REBECCA project	1638188948376_dt30lk	30 Nov 2021 17:21:01	new
#3	https://rebeccaproject.eu.teamwork.com/#/projects/437581/messages	REBECCA - Messages - REBECCA project	1638188948376_dt30lk	30 Nov 2021 17:20:57	new
#4	https://rebeccaproject.eu.teamwork.com/#/messages/625514	REBECCA - Messages - REBECCA project	1638188948376_dt30lk	30 Nov 2021 17:20:54	new

#	Query	Session ID	Timestamp	Tab
#1	mydata+e+timologio	1638188948376_dt30lk	30 Nov 2021 17:06:51	new
#2	Verification+of+News+Video+Content%3A+Findings+from+a+Study+of+Journalism+Students	1638188948376_dt30lk	30 Nov 2021 16:55:59	new
#3	visula+paradigm+change+font	1638188948376_dt30lk	30 Nov 2021 15:03:28	existing



March 28, 2022



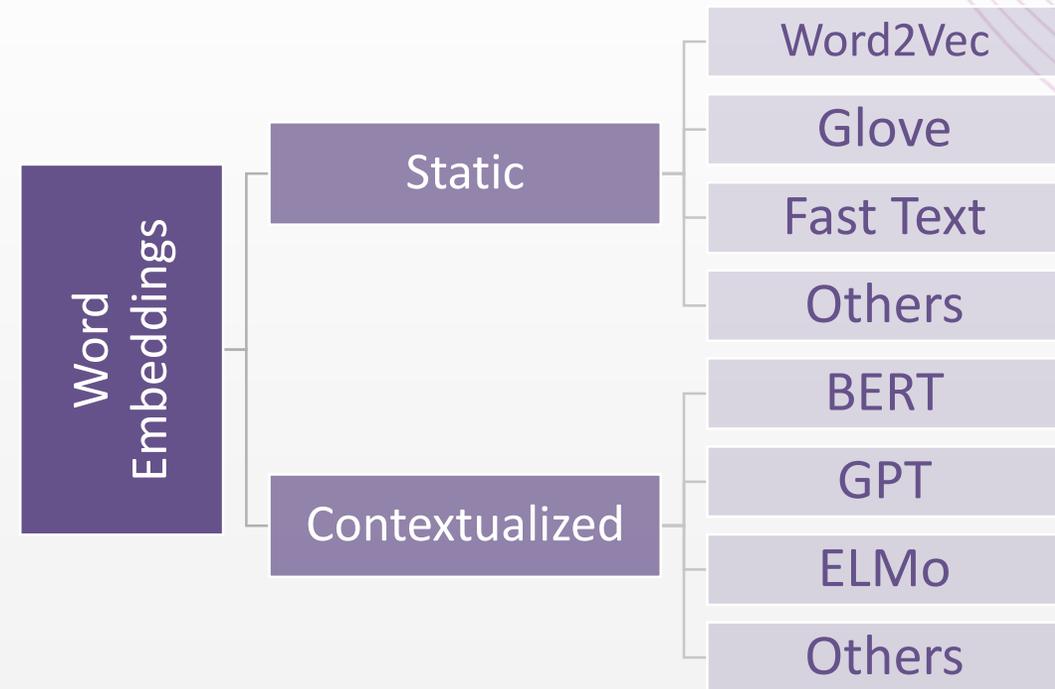
This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# Pre-processing

## Traditional Features

- Presence based, frequency based, syntax
- Lexicon usage is popular
  - common lexicons:
    - [WordNet](#), [DepecheMood](#), [NRC](#), etc.

## Word/Sentence representation



March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# Related work

- In the literature, classical machine learning models as well as deep learning models have been used with good results
- Pre-trained models, such as BERT, show potential
- No single model outperforms others in the literature, highly depends on pre-processing and dataset



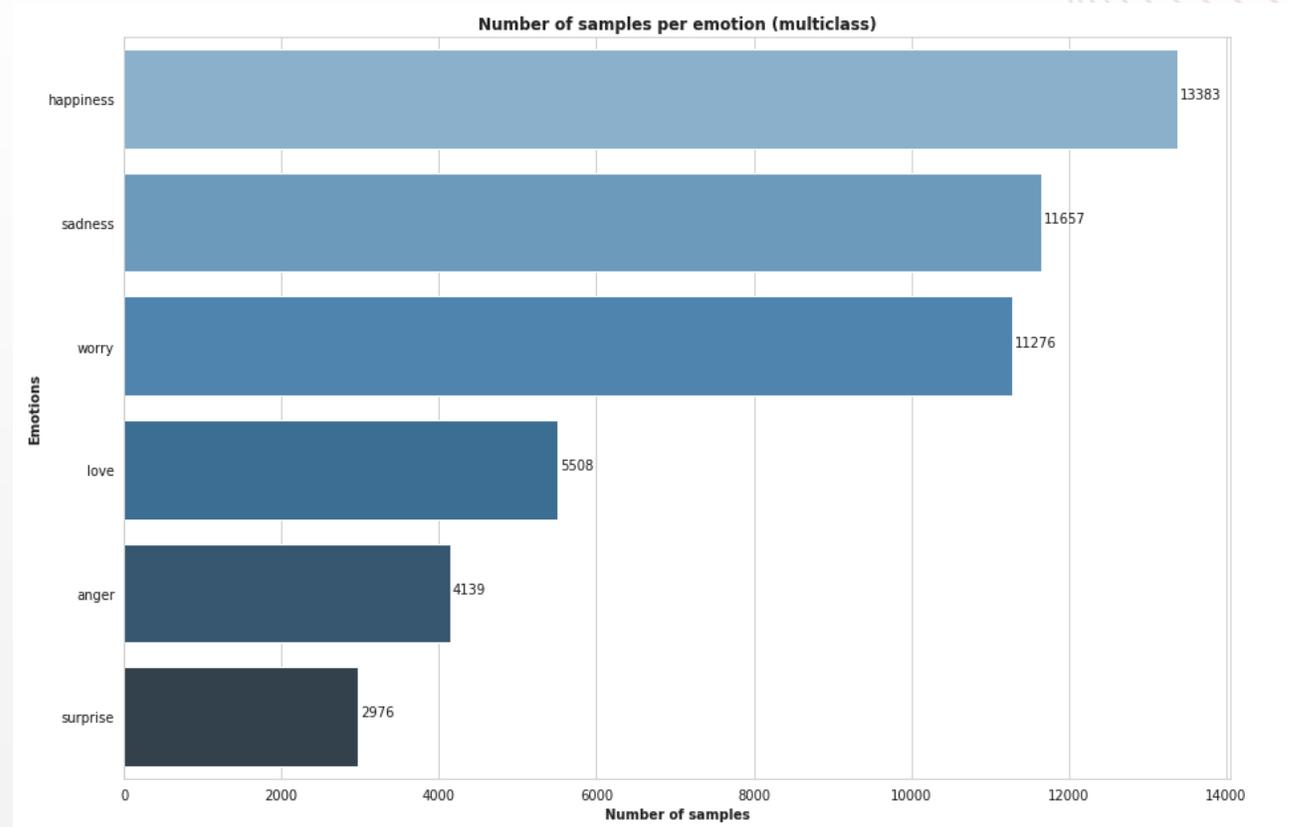
March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# Our preliminary work

- Multiclass problem
  - Created a dataset by merging other datasets – enriched some classes
  - Total: 49K instances
  - Emotions: sadness, worry, surprise, love, happiness, anger



March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# Our preliminary work



- Multiclass for merged dataset (6 classes):
  - Created baseline models – NB, LR, RF, Linear SVM
  - Implemented CNN, LSTM and Bi-LSTM
  - Fine tuned BERT
  - BERT together with an LSTM layer

Model	Accuracy	F1 macro	Macro Recall	Macro Precision
Naive Bayes	49.39%	0.2940	0.3311	0.5399
Logistic Regression	57.68%	0.4837	0.6443	0.6410
Random Forest	53.06%	0.4554	0.4320	0.5971
Linear SVM	60.69%	0.5586	0.5379	0.6142
CNN	64.87%	0.6146	0.6012	0.6719
bi-LSTM	66.42%	0.6250	0.6102	0.6929
LSTM	67.10%	0.6398	0.6209	0.6954
BERT	67.29%	0.6406	0.6400	0.6400
BERT + LSTM	67.84%	0.6522	0.6400	0.6700



March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# Issues and Challenges

## Data gathering

- Privacy issues – ethical issues – invasiveness into private life
- Low or no use of the Internet, especially by older women
- Annotation difficulties

## Emotional well-being identification

- Many confounding factors – no clear association between online behavior and emotional well-being
- Several languages
- Domain adaptation for cancer survivors
- Short text is limiting



March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# Future of the project

- Emotional well-being identification additions and improvements
  - text analysis in order to detect depression or anxiety signs
  - use data from online activity patterns as well as behavioral indicators
- Evaluation of the model and the generated data in three clinical studies – they aim to *improve patient management and patient outcomes* (eg. QoL)
  - The clinical sites are Stavanger (Norway), Stockholm (Sweden) and Valencia (Spain)

CD  
CD



March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

# Thank you!



Georgia Pantalona, Lazaros Apostolidis,  
Filareti Tsalakanidou, Symeon  
Papadopoulos

Images:flaticon.com and vecteezy.com



March 28, 2022



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.