



AI@Smile

Artificial intelligence and sustainable educational tools
for the prevention of teenage psychological disorders

Work Package 4: AI Chatbot

Activity WP4.2 Data-mining

Data-mining methodology guide



Co-funded by
the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Project acronym:	AI@SMILE
Project full title:	Artificial intelligence and sustainable educational tools for the prevention of teenage psychological disorders
Grant agreement no.:	KA220-NI-23-36-161162
Responsible partner for deliverable:	University of Novi Sad, Serbia
Contributing partners:	All partners/Whole consortium
Author(s):	Bojana Dinić, Milica Solarević, Miroslav Vujičić
Distribution level:	
Total number of pages:	14 pages
Version:	
Language	English
Reviewed by:	
Status:	
Delivery date:	

Version control

Number	Date	Description
V1.	23.07.2024.	First draft

Contents

Abbreviations and Acronyms	3
1. Overview of the Work Package 4: AI Chatbot	4
2. Data-mining methodology	5
2.1 Research instrument: YIEPS	5
2.2 Open-ended questions	6
2.3 Ethical approval.....	6
2.3.1 Informed consent	7
2.4 Instrument's distribution, sampling and data collection.....	8
3. Annexes	10
3.1 Annex 1: Instruments	10
3.2 Annex 2: Informed Consent for STUDENTS who don't need parental consent.....	15
3.3 Annex 3: Informed Consent for STUDENTS with PARENTAL/LEGAL GUARDIAN'S CONSENT	17

Abbreviations and Acronyms

Abbreviation	Full name
Acronym	Full name
EACEA	Education, Audiovisual and Culture Executive Agency
EC	European Commission
EU	European Union
GA	Grant Agreement
HEI	Higher Education Institution
ICT	Information and Communication Technologies
PC	Project Coordinator
WP	Work Package
AI	Artificial intelligence

1. Overview of the Work Package 4: AI Chatbot

The main specific objectives of WP4 aiming at creating an AI based tool for teenagers mental health disorders' detection, as follows:

- Early Detection: develop a tool that can identify early signs of mental health disorders in teenagers, allowing for prompt intervention and treatment. This will contribute in raising awareness in both teenagers and adults about situations of vulnerability and will allow them to take immediate therapeutic actions to reduce symptoms of anxiety, depression; for example, through the introduction of nature-based and sustainable activities as preventive and therapeutic measures for mental health and well-being.
- Preserve the child's privacy, which will contribute to the general objective of the project that foresees creating a safe and supportive environment in which the mental health stigma is removed and the teenagers can freely express themselves and can ask for support, if they need it. The tool will be clearly designed keeping privacy in mind, ensuring that the data collected from teenagers is kept confidential and secure (refer to WP2).
- Accuracy: the tool will be developed in an accurate way (data-mining will be crucial for this purpose), allowing teachers, educators, parents, and caregivers to detect mental health disorders in accessing reliable information that can be used to inform clinical decisions. This specific objective will contribute to increasing education and awareness among adults about early signs of mental health disorders, self-harm, and suicidal ideation in teenagers. The tool will be developed in a user-friendly way, making it easy for both teenagers and healthcare professionals to use and to provide personalized "treatment" suggestions, for example, suggesting and referring to educational contents of WP3.

The main result of this WP will be the innovative development and introduction of an AI toolkit to detect early signs of mental health deterioration in teenagers. Application of AI to this field, especially targeting minors, can be considered a very ambitious action and will position AI@Smile as a pilot on this ground.

In this work package BIRA will play the major role, especially it's IT technicians, however, the active participation of all the partners will be required, each one performing a specific task and function, assigned according to their professional expertise.

1. Needs analysis / Desk Research (UA leads -all Universities take part)
- 2. Data-mining (UNS leads - all the partners take part)**
3. Labeling / Data warehouse (UA Leads - All the Partners take part)
4. Design (BIRA)
5. Model training (BIRA)
6. Testing (Oldenburg Leads - all the partners take part)

The main results of this WP will be:

- 1) A report concerning the main findings of the need analysis, held by the Universities in the targeted schools through questionnaires;
- 2) A collection of the data needed for the machine learning preparation, resulting from the data-mining process held in the targeted schools;**
- 3) A document from each school with labeling of the data collected;
- 4) The launch of the first version of the AI Chatbot;
- 5) A report of the model training realized;

- 6) A report of the testing realized in the targeted schools.

The activities of this WP will provide a robust sample size that allows for a comprehensive analysis of the data. The researchers will gather data using surveys and questionnaires, and the resulting data will be used to identify patterns and trends that, besides allowing the training of the AI Chatbot (machine learning) to provide more accurate information, can be used for the development of effective strategies to combat mental health deterioration.

2. Data-mining methodology

Name of the activity: **Data-mining**

Duration of the activity: **19.02.2024 – 17.11.2024**

Deliverable/Result: **Collection of the data needed for the machine learning preparation**

Activity 2 within WP4 is Data-mining activity, led by the **University of Novi Sad**. This activity will be based on collecting the data for the machine learning process and developing AI Chatbot. All the schools involved in this project must participate in it, since they will have the crucial role of providing information to feed the AI tool with data that will be collected through ad-hoc questions and validated instrument selected by the collaboration among the Universities. The rest of the partners will also be in charge of disseminating the aforementioned instruments.

To ensure the monitoring and transparency of this activity, a designated folder was created at Google Drive, as [WP4.2 Data-mining \(UNS\)](#). In this folder, you will find all the files needed for this activity. Each country has its own subsection, to ensure proper monitoring of the work progress separately and timely.

2.1 Research instrument: YIEPS

We chose the **Youth Internalizing + Externalizing Problems Screeners (YIEPS)**. The YIEPS is developed as a pair of self-report, evidence-based rating scales for assessing youths' broad mental health problems. We choose it as our main tool because: it is a short, screening instrument for assessing youths' broad mental health problems; it is in the open domain; it is adequate for adolescents (our target group); and it is intended for use in school mental health research.

The YIEPS is composed of two subscales: (1) Internalizing + (2) Externalizing. Subscale scores can be used as standalone measures of broad mental health problems.

License and use: The YIEPS is a “free cultural work,” licensed under a Creative Commons–Attribution 4.0 International license (CC BY 4.0). There is no cost for using the YIEPS and no need to obtain permission prior to use. You can learn more about the terms of the CC BY 4.0 licensing agreement here: <https://creativecommons.org/licenses/by/4.0/>.

Although it is a widely used instrument, it has not been adapted to all the languages in which our research will be conducted, so adaptation into some languages will be necessary. Details about the instrument could be found at <https://osf.io/ets7c>. Before translating and adapting

the YIEPS to the local language, please check if there is an existing adaptation that you can use.

Reference for this free version of the YIEPS with accompanying User Guide is: Renshaw, T. L. (2020). *Youth Internalizing + Externalizing Problems Screeners(YIEPS): Measure and user guide*. Open Science Framework. Available at <https://osf.io/ets7c>

2.2 Open-ended questions

To ensure variability in the data, questions about socio-demographic characteristics open-ended questions were included at the beginning of the data collection process. It is important to include these questions initially so that participants can freely express themselves, thereby setting the framework for their responses on the YIEPS. Since we already have a scoring key for YIEPS, we can explore differences in the responses to the open-ended questions between those who score high and low on the YIEPS scales.

We should keep in mind that we did not include an open-ended question for each possible symptom of mental health, as provided in the instructions, because the main symptoms are already covered in the YIEPS. Instead, open-ended questions are focused on asking about feelings, thoughts, activities, and interpersonal relationships with peers (see Annex 1 for these four sets of questions). We provided examples and possible directions for responses to ensure that the answers remain relevant to the specific questions.

For example, one of the questions is, “Can you describe in several sentences your main activities during the day, except going to school. How have you usually spent your free time in the past month?”. By analyzing the relationship between YIEPS responses and answers to these open-ended questions, we can identify potential indicators (in this case, types of activities) associated with mental health issues. For instance, if a participant mentions engaging in extracurricular activities, attending training sessions, spending time with peers, or reading, and if these answers were associated with lower levels of mental health issues based on YIEPS, these activities could be categorized as adaptive behaviors. Conversely, if a participant's answers include sleeping, watching TV, or spending excessive time on the internet, and these answers were linked to high YIEPS scores, these activities could be categorized as potential indicators connected to mental health problems.

2.3 Ethical approval

Before starting the research, ethical approval should be obtained.

HERE WILL BE DESCRIBED AND LINKED ETHICAL APPROVAL OF THE UNSPMF ETHICAL BOARD

Please contact your local Institutional Ethical Board to check whether you need separate ethical approval or it is enough to use ours. Usually, the local Institutional Ethical Board will ask for local approval.

If you need it, here is the information for submission of research to Ethical Board:

The purpose of the research: The purpose of this research is to obtain text data from adolescents about their potential mental health issues, which will be used as training data for developing an app for recognizing and mental health issues in adolescents.

Sample recruiting: The required sample size is 400 secondary school students (or 200, see 2.4). Before conducting the research, approval from school management will be obtained (in written or oral form). After this approval, researchers from each country will contact the teachers, psychological-pedagogical staff, or individuals recommended by the school management to organize the distribution of instruments. In agreement with them, informed consent for parents to give the consent that their child participates in the research will be distributed to parents via usual way of communicating with parents (e-mail or Viber groups, meetings, etc.). Signed informed consent will be uploaded in the provided Google Drive folder. Only participants who parent and themselves give consent will be included in the research. To them, a link to the instruments will be sent. Data will be collected via the Google Form platform.

Usually, the instruments will also need to be submitted to ethical approval. In that case, provide a translated version of the open-ended questions and YIEPS. Please, upload the approval in the folder dedicated to your country/institution.

2.3.1 Informed consent

All participants need to give consent for participating in the research. Since some of the participants will be minors, one of their parents must approve their participation in the research. Please note that each country has its own rules regarding the minimum age of participants, which do not need to have approval from parents (16 or 18). In any case, parents need to be informed about the research.

Having that in mind, as well as the sensitivity of the topic and ensuring the unique procedure, for this activity, **two different informed consent was created** (check the Annexes) **for all respondents**. For minor students, their parents or legal guardians will need to check that they understand the information, give consent, and fill in their name and the name of the child. Students who don't need the consent according to national regulations will sign the Consent by themselves. Only those participants whose parent gives consent (if needed) and who also give consent themselves should be included in the research, and their data can then be recorded.

All signed consent forms will be uploaded in a designated **folder in Google drive** via <https://drive.google.com/drive/folders/1pB48sLn4v7PCypM-0jF8ahJTrOQALDL1>. Each school/country has its own designated folder to track the number of collected consents and completed questionnaires. To upload consent forms, please use the folder designated to your country. If you find it useful, you can create subfolders to separate classes.

The English version of the Informed consent is provided within the designated folder in Google Drive, as well as **Annex 1** at the end of this document, and all the countries should translate it to their own national language, before distribution, if needed.

2.4 Instrument's distribution, sampling and data collection

This activity will be based on collecting the data through open-ended questions and YIEPS instrument. All the schools involved in this project must participate in it, since they will have the crucial role of providing information to feed the AI tool. The rest of the partners will also be in charge of disseminating the aforementioned questionnaires.

This research activity is cross-national. In **each country that has associated school partner (Serbia, Spain, Hungary, Türkiye)**, a sample of **400 secondary school students per country** will be selected from local schools, some of which are project partners. Since the number of students from some schools is less than the targeted number, each country has to provide additional students to complete the instruments, contacting schools from local networks, community or abroad, to ensure a total of 400 students. Following schools are involved in this activity:

- IES DOCTOR BALMIS, Spain
- SAMSUN ANADOLU LİSESI, Türkiye
- Váci SZC Selye János Egészségügyi Technikum, Hungary
- Racunarska gimnazija Smart Novi Sad, Serbia

Since all partners have to participate in this activity, countries that don't have associated schools in this project **(Germany and Greece) will gather 200 students per country**, following the same procedure, explained within this Methodology guide. This is the minimum of a required sample per country in order to check the cross-cultural validity of the YIEPS, which is important for score comparison.

Summarizing, this means as follows:

Number of students per country:

SERBIA, SPAIN, HUNGARY, TÜRKIYE: 400 students

GERMANY AND GREECE: 200 students

TOTAL number of students/respondents: **2000 students**

Before conducting the research, approval from school management will be obtained, written or oral. After this approval, researchers from each country will contact the teachers, psychological,-pedagogical staff, or individuals recommended by the school management to organize data collection.

Data will be collected through a **specially designed Google Forms translated and adapted INSTRUMENTS to all partners' languages**, which can be used simultaneously on various devices (mobiles, tablets, laptops, desktop computers). Participation in the research could be organized in schools during class, as a recommended way, or the link could be sent to participants.

Final versions of the instruments can be accessed through designated [Data collection](#) Google folder, or following the links provided below for versions in national languages. [English version](#) of the questionnaires was provided as Google forms template for the translation process and can be accessed in the mentioned folder.

[Serbian version](#)

[Spanish version](#)

[Hungarian version](#)

[Turkish version](#)

[German version](#)

[Greek version](#)

3. Annexes

3.1 Annex 1: Instruments

Open-ended questions

- Can you describe in several sentences how you've been feeling in the past month? Please share any specific emotions you've experienced, such as happiness, sadness, frustration, anxiety, etc.
- Whether your emotional state remained the same or changed in the past month? Could you give me more details?
- Can you describe in several sentences what your thoughts are in the past month? Whether you were occupied with some thoughts?
- Can you describe in several sentences your main activities during the day, except going to school? How do you usually spend your free time in the past month?
- Can you describe your interpersonal relationships with your peers over the past month in a few sentences? For example, consider whether you have been getting along and spending quality time together, or if there have been times when you felt angry at each other or felt upset by their behavior.

YIEPS

The following statements are about how you can feel, think, and act. Read each statement and choose one of the four that best describes how you feel and think in the PAST MONTH.

1. Choose one of these four statements that best describes how you feel, think and work in the PAST MONTH?
 - I almost never feel nervous or afraid.
 - I sometimes feel nervous or afraid.
 - I often feel nervous or afraid.
 - I almost always feel nervous or afraid.
2. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
 - I almost never feel very tired or drained of energy.
 - I sometimes feel very tired or drained of energy.
 - I often feel very tired or drained of energy.
 - I almost always feel very tired or drained of energy.
3. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
 - I almost never find it hard to relax and settledown.
 - I sometime find it hard to relax and settledown.
 - I often find it hard to relax and settledown.
 - I almost always find it hard to relax and settledown.

4. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
 - I almost never get bothered by things that didn't bother me before.
 - I sometimes get bothered by things that didn't bother me before.
 - I often get bothered by things that didn't bother me before.
 - I almost always get bothered by things that didn't bother me before.

5. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
 - I almost never have uncomfortable and tense feelings in my body.
 - I sometimes have uncomfortable and tense feelings in my body.
 - I often have uncomfortable and tense feelings in my body.
 - I almost always have uncomfortable and tense feelings in my body.

6. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
 - I almost never feel moody or grumpy.
 - I sometimes feel moody or grumpy.
 - I often feel moody or grumpy.
 - I almost always feel moody or grumpy.

7. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
 - I almost never feel like I'm going to panic or think I might lose control.
 - I sometimes feel like I'm going to panic or think I might lose control.
 - I often feel like I'm going to panic or think I might lose control.
 - I almost always feel like I'm going to panic or think I might lose control.

8. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
 - I almost never feel like I do not really enjoy doing anything anymore.
 - I sometimes feel like I do not really enjoy doing anything anymore.
 - I often feel like I do not really enjoy doing anything anymore.
 - I almost always feel like I do not really enjoy doing anything anymore.

9. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
 - I almost never feel worthless or lonely when I'm around other people.
 - I sometimes feel worthless or lonely when I'm around other people.
 - I often feel worthless or lonely when I'm around other people.
 - I almost always feel worthless or lonely when I'm around other people.

10. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
 - I almost never have headaches, stomachaches, or other pains.
 - I sometimes have headaches, stomachaches, or other pains.
 - I often have headaches, stomachaches, or other pains.
 - I almost always have headaches, stomachaches, or other pains.

11. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
- I almost never lose my temper and get angry with other people.
 - I sometimes lose my temper and get angry with other people.
 - I often lose my temper and get angry with other people.
 - I almost always lose my temper and get angry with other people.
12. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
- I almost never have a hard time sitting still when other people want me to.
 - I sometimes have a hard time sitting still when other people want me to.
 - I often have a hard time sitting still when other people want me to.
 - I almost always have a hard time sitting still when other people want me to.
13. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
- I almost never fight and argue with other people.
 - I sometimes fight and argue with other people.
 - I often fight and argue with other people.
 - I almost always fight and argue with other people.
14. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
- I almost never break rules whenever I feel like it.
 - I sometimes break rules whenever I feel like it.
 - I often break rules whenever I feel like it.
 - I almost always break rules whenever I feel like it.
15. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
- I almost never talk a lot and interrupt others when they are talking.
 - I sometimes talk a lot and interrupt others when they are talking.
 - I often talk a lot and interrupt others when they are talking.
 - I almost always talk a lot and interrupt others when they are talking.
16. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
- I almost never say or do mean things to hurt other people.
 - I sometimes say or do mean things to hurt other people.
 - I often say or do mean things to hurt other people.
 - I almost always say or do mean things to hurt other people.
17. Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
- I almost never have a hard time focusing on things that are important.
 - I sometimes have a hard time focusing on things that are important.
 - I often have a hard time focusing on things that are important.
 - I almost always have a hard time focusing on things that are important.

- 18.** Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
- I almost never like to annoy people or make them upset.
 - I sometimes like to annoy people or make them upset.
 - I often like to annoy people or make them upset.
 - I almost always like to annoy people or make them upset.
- 19.** Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
- I almost never get distracted by the little things happening around me.
 - I sometimes get distracted by the little things happening around me.
 - I often get distracted by the little things happening around me.
 - I almost always get distracted by the little things happening around me.
- 20.** Which one of these four statements best describes how you feel, think and work in the PAST MONTH?
- I almost never choose not to follow directions and don't listen to adults.
 - I sometimes choose not to follow directions and don't listen to adults.
 - I often choose not to follow directions and don't listen to adults.
 - I almost always choose not to follow directions and don't listen to adults.

3.2 Annex 2: Informed Consent for STUDENTS who don't need parental consent

Informed Consent for STUDENTS who don't need parental consent

Purpose:

The main goal of this research is to collect the data for developing a chatbot that will aim at prevention and detection of mental health issues among adolescents. This will contribute to raising awareness in both teenagers and adults about situations of vulnerability and will aim at promoting teenagers' mental wellbeing.

The research was supported by the **Erasmus+ Project AI@Smile (KA220-NI-23-36-161162)**.

What participation in the research involves:

Participation is voluntary, anonymous, and will not require any personal information. All data will be analyzed at the group level, and there will be no possibility of connecting your specific answers to you. The collected data will be used exclusively for the purpose of developing the described application. You will be asked to provide some demographic data about yourself, namely your gender and age. Then, you will be asked more questions and rate feelings and thoughts on various aspects. This information will help us finalize the results as explained in this form.

Duration: The survey takes around 15 minutes.

Potential risks: If any questions cause you discomfort or concern, you can call the following numbers for support: [add national hotlines or websites for adolescent support, or if you have an agreement with psychological-pedagogical staff from the school, their contact information could be provided].

Confidentiality Measures:

The participation is anonymous and only members of the research team will have access to the data collected. Data will be securely stored and used only for the purpose of facilitating project activities. After the end collection data will be downloaded to the computers of project participants only and deleted from the platform, and then it will be locked and encrypted only on the computers of the project participants.

Right to withdraw from the research:

If you encounter any type of discomfort, you can stop at any time without prior justification. You can withdraw from the study at any time without giving explanations and with no negative consequences.

Following Article 14 of the European General Data Protection Regulation, this signed consent form will be kept by the AI@Smile project consortium to prove your consent.

Contact person

If you have any questions or concerns about the informed consent process or the background information form regarding this activity, please do not hesitate to contact [add Name and Surname of the contact person from your institution; add email address of contact person] on behalf of AI@Smile consortium and ask for more information.

STUDENT'S CONSENT

- I have read the information about the research project, and I have had the opportunity to ask questions, which have been answered to my satisfaction.
- I agree to take part in this research, and I have received a copy of this consent form.

Full name and surname: _____

Signature: _____

Date: _____

3.3 Annex 3: Informed Consent for STUDENTS with PARENTAL/LEGAL GUARDIAN'S CONSENT

Informed Consent for STUDENTS with PARENTAL/LEGAL GUARDIAN'S CONSENT

Purpose:

The main goal of this research is to collect the data for developing a chatbot that will aim at prevention and detection of mental health issues among adolescents. This will contribute in raising awareness in both teenagers and adults about situations of vulnerability and will aim at promoting teenagers' mental wellbeing.

The research was supported by the **Erasmus+ Project AI@Smile** (KA220-NI-23-36-161162).

What participation in the research involves:

Participation is voluntary, anonymous, and will not require any personal information. All data will be analyzed at the group level, and there will be no possibility of connecting your specific answers to you. The collected data will be used exclusively for the purpose of developing the described application. You will be asked to provide some demographic data about yourself, namely your gender and age. Then, you will be asked more questions and rate feelings and thoughts on various aspects. This information will help us finalize the results as explained in this form.

Duration: The survey takes around 15 minutes.

Potential risks: If any questions cause you discomfort or concern, you can call the following numbers for support: [add national hotlines or websites for adolescent support, or if you have an agreement with psychological-pedagogical staff from the school, their contact information could be provided].

Confidentiality Measures:

The participation is anonymous, and only members of the research team will have access to the data collected. Data will be securely stored and used only for the purpose of facilitating project activities. After the end of the collection, data will be downloaded to the computers of project participants only and deleted from the platform, and then it will be locked and encrypted only on the computers of the project participants.

Right to withdraw from the research:

If you encounter any type of discomfort, you can stop at any time without prior justification. You can withdraw from the study at any time without giving explanations and with no negative consequences.

Following Article 14 of the European General Data Protection Regulation, this signed consent form will be kept by the AI@Smile project consortium to prove your consent.

Contact person

If you have any questions or concerns about the informed consent process or the background information form regarding this activity, please do not hesitate to contact [add Name and Surname of the contact person from your institution; add email address of contact person] on behalf of AI@Smile consortium and ask for more information.

STUDENT'S CONSENT

- I have read the information about the research project, and I have had the opportunity to ask questions, which have been answered to my satisfaction.
- I agree to take part in this research, and I have received a copy of this consent form.

Full name and surname: _____

Signature: _____

Date: _____

PARENTAL/LEGAL GUARDIAN'S CONSENT

- I have read the information about the research project, and I have had the opportunity to ask questions, which have been answered to my satisfaction.
- I give consent for my child to participate in the research, and I have received a copy of this consent form.

Full name and surname: _____

Signature _____

Date: _____