Journal of Responsible Technology 22 (2025) 100119

Contents lists available at ScienceDirect

Journal of Responsible Technology

journal homepage: www.sciencedirect.com/journal/journal-of-responsible-technology

Research Article

Decision-making on an AI-supported youth mental health app: A multilogue among ethicists, social scientists, AI-researchers, biomedical engineers, young experiential experts, and psychiatrists

Dorothee Horstkötter [a,*], Mariël Kanne [b], Simona Karbouniaris [c], Noussair Lazrak [d], Maria Bulgheroni [e], Ella Sheltawy [f], Laura Giani [e], Margherita La Gamba [e], Esmeralda Ruiz Pujadas [d], Marina Camacho [d], Finty Royle [f], Irene Baggetto [f], Sinan Gülöksüz [g], Bart Rutten [g,h], Jim van Os [i]

[a] *Department of Health, Ethics and Society, Mental Health and Neuroscience Research Institute, Faculty of Health Medicine and Life Sciences, Maastricht University, Postbox 616, 6200 MD Maastricht, The Netherlands*
[b] *Research Group Living and Wellbeing, Utrecht University of Applied Science, Utrecht, The Netherlands*
[c] *Centre for Social Innovation, Research Group Participation, Care & Support, Utrecht University of Applied Sciences, Utrecht, The Netherlands*
[d] *Department of Mathematics and Computer Science, BCN-AIM lab, University of Barcelona, Barcelona, Spain*
[e] *Ab.Acus, Milano, Italy*
[f] *Euro Youth Mental Health, CIC 21 Ninesprings Way, Hitchin, SG4 9NN United Kingdom*
[g] *Department of Psychiatry and Neuropsychology, Mental Health and Neuroscience Research Institute, Maastricht University, Maastricht, The Netherlands*
[h] *Maastricht University Medical Centre, Maastricht, The Netherlands*
[i] *Utrecht Medical Center, Utrecht, The Netherlands*

ARTICLE INFO

Keywords:
AI-supported mental health care
Autonomy
Biomedicine
Co-creation
Digital mental health
Epistemic pluralism
Ethics
EU-funding
Expertise by experience
Human-decision making
Judicial regulations
Multilogue
Transdisciplinary development
Youth mental health care
Artificial intelligence

ABSTRACT

This article explores the decision-making processes in the ongoing development of an AI-supported youth mental health app. Document analysis reveals decisions taken during the grant proposal and funding phase and reflects upon reasons *why* AI is incorporated in innovative youth mental health care. An innovative multilogue among the transdisciplinary team of researchers, covering ethicists, social scientists, AI-experts, biomedical engineers, young experts by experience, and psychiatrists points out *which* decisions are taken *how*. This covers i) the role of a biomedical and exposomic understanding of psychiatry as compared to a phenomenological and experiential perspective, ii) the impact and limits of AI-co-creation by young experts by experience and mental health experts, and iii) the different perspectives regarding the impact of AI on autonomy, empowerment and human relationships. The multilogue does not merely highlight different steps taken during human decision-making in AI-development, it also raises awareness about the many complexities, and sometimes contradictions, when engaging in transdisciplinary work, and it points towards ethical challenges of digitalized youth mental health care.

> "Not everything that can be counted counts and not everything that counts can be counted" (attributed to Albert Einstein)

* Corresponding author at: Department of Health, Ethics and Society, Mental Health and Neuroscience Research Institute, Faculty of Health Medicine and Life Sciences, Maastricht University, Postbox 616, 6200 MD Maastricht, The Netherlands.

E-mail addresses: d.horstkoetter@maastrichtuniversity.nl (D. Horstkötter), mariel.kanne@hu.nl (M. Kanne), simona.karbouniaris@hu.nl (S. Karbouniaris), noussair.lazrak@ub.edu (N. Lazrak), mariabulgheroni@ab-acus.eu (M. Bulgheroni), ellas@eymh.org (E. Sheltawy), lauragiani@ab-acus.eu (L. Giani), margheritalagamba@ab-acus.eu (M. La Gamba), esmeralda.ruiz@ub.edu (E. Ruiz Pujadas), marina.camacho@ub.edu (M. Camacho), sinan.guloksuz@maastrichtuniversity.nl (S. Gülöksüz), b.rutten@maastrichtuniversity.nl (B. Rutten), j.j.vanOs-2@umcutrecht.nl (J. van Os).

https://doi.org/10.1016/j.jrt.2025.100119

Available online 18 April 2025

1. Introduction

This article aims to explore decision-making processes in the development of an AI-supported youth mental health app, with a focus on ethical considerations, data privacy, and the integration of multidisciplinary perspectives. In the past decades, youth mental health and youth mental health care have gained increasing attention in the public sphere, among young people themselves, and in academic research discourses. The background of this development is that more and more young people report symptoms such as depression, anxiety, and other mental health troubles and seek help and support. According to some, we are currently being confronted with a 'youth mental health crisis' (Aftab & Druss, 2023; European Commission, 2023; Nature Mental Health, 2023). This is particularly problematic, given that the mental health of young people has a large impact on mental health burden and quality of life across people's whole lifespan. To tackle these challenges, there is increasing attention for the prospects of digital mental health technologies. Smartphone-based mental health apps and digital chatbots promise to be rather easily accessible, well accepted by a young 'digitally native' population, and effective in addressing a wide variety of trans-syndromal mental health issues (Amir & Torous, 2017; Daley, Hungerbuehler, Cavanagh, Claro, Swinton & Kapps, 2020; Smith et al., 2023). Computational tools that are supported by artificial intelligence and machine learning are developing at a fast pace (Garcia-Ceja, Riegler, Nordgreen, Jakobsen, Oedegaard & Tørresen, 2018; Le Glaz et al., 2021).

1.1. The digital turn in youth mental health

This digital and AI-turn in youth mental health care currently meets research approaches on gene-environment interaction that emphatically points towards the complex interrelationships between environmental, genetic and (neuro)biological features as being crucial for the understanding of the onset and pathways of mental health trajectories (Klengel & Binder, 2015; Musci, Augustinavicius & Volk, 2019; Renzi et al., 2018). According to this view, in order for mental ill health to be manifested in a young person, a variety of social and environmental stressors, covering among others adverse childhood experiences, lack of social support, economic hardships, substance abuse, or the downsides of social media exposure and digitalization interact with distinct molecular and neurobiological underpinnings of mental health.

Scientific research on these various contributors to mental health issues has generated large amounts, and different types, of data. The hope is that the analysis of multi-cohort data may provide new insights into the gene-environment interactions underlying actual youth mental health. To maximize the usage of these existing data increasingly AI-driven methods are being used to identify trajectories of mental health previously unrecognized. Moreover, AI-supported analysis of existing data is supposed to be able to innovate in the mental health sphere, where distinct psychiatric diagnoses are increasingly debated and the focus is rather on transdiagnostic syndromes (Fusar-Poli et al., 2019). A data-driven analysis of comprehensive datasets promises the identification of new subgroups of mental health outcomes that might operate independently from diagnostic categories. At the same time, ongoing measurement and assessment of real-time physiological and social behavioral parameters gained via passive digital phenotyping using young people's smartphones, add further information to an AI-driven analysis of risk prediction and treatment advice on an individual level (Moura, Teles, Viana, Marques, Coutinho & Silva, 2023).

1.2. Ethical considerations on AI in youth mental health

While AI-modeling and AI-supported analysis in mental health care are gaining momentum, there are increasing ethical concerns as to the implications of these developments. These largely cover issues regarding privacy and data-protection, concerns such as implementing end-to-end encryption, anonymization of user data, and compliance with the General Data Protection Regulation (GDPR) to safeguard patient information. Trust in therapeutic relationships or extensive continuous surveillance raises questions about how AI tools might impact these relationships. AI-supported recommendations could strengthen the therapeutic alliance by providing additional insights, but potentially also undermine trust if perceived as impersonal or overly invasive by patients (Grabb, Lamparth & Vasan, 2024; Martinez-Martin, 2021, 2018).

A particular ethical concern covers the question of the role of human agency and human decisions in the various steps of developing AI-mediated health and mental health care. This question arises at the very end of the process when decisions about appropriate prediction, diagnosis or therapeutic advice are being taken, but also at the very beginning of the design of algorithms and (multi)models and when decisions must be taken that are often inherently normative. According to standing ethical AI-principles (c.f. for an overview Siala & Wang, 2022), human-centeredness is crucial. This principle requires that a human being remains integral to medical decision-making. In practice, this means that AI-generated recommendations are to be used as supplementary tools that clinicians review and interpret before finalizing any medical decisions. To ensure the preservation of human judgment and empathy, current EU-law grants patients a right to human oversight. Still, it remains highly debated, what this right entails and how human and artificial systems can collaborate, particularly in cases where they come to different conclusions (Grote & Berens, 2020, 2022; Ho, 2019; Salloch & Eriksen, 2024). At the early stages, it is not so much AI-ethical principles that get invoked to support ethical decision making, but rather, the 'ethics by design' framework should ensure that justifiable normative decisions are taken when data-processing models are given shape, for example such that outcomes are rather unbiased regarding, for example, gender, age, or ethnicity (Brey & Dainow, 2023; d'Aquin, Troullinou, O'Connor, Cullen, Faller & Holden, 2018).

1.3. Human-decision making on digital AI-tools in youth mental health

Much less attention has been given to the background and context in which decisions on AI are taken. The decision to even develop an AI-tool and to test and use it for a medical purpose does not appear out of nowhere. This holds particularly in the context of mental health care and youth mental health care, where in addition to a complex set of measurable and testable molecular, neurobiological and environmental features, a qualitative experiential dimension has to be taken into account. In this sense, the role of AI in mental health care is gaining increasing attention. Mental health care is a new, but still underexplored, area of application for quantifying AI-tools. (Melissa, Katrina & Daniel, 2023; Starke, De Clercq, Borgwardt & Elger, 2021). However, given that mental health has also an inherent phenomenological and experiential dimension, it is crucial to anticipate the potential incompatibility of the phenomenological with the arithmetical (Aho, 2008).

The aim of the current contribution is to point out how and which humans are in the loop in the background making decisions in a research project that develops and investigates AI-tools in the context of innovative youth mental health care. We engage in a transdisciplinary multilogue and present and discuss the various steps in the human-led decision-making processes taken in the study. We put particular focus on the development of an AI-supported mental health app that will be tested in a subsequent clinical trial with help-seeking young people aged 12 to 24 years to be carried out in a subsequent research phase.

2. The Youth-GEMs consortium and the app

To better understand the highly complex interplay between environmental, biomedical and experiential features of young people's trajectories in mental health and to use these insights for the development

of AI-supported modes of prevention and intervention is the proposed aim of the EU-Horizon funded consortium entitled 'Youth-GEMs, Gene Environment Interactions in Mental Health Trajectories of Youth". This European consortium combines fundamental neuroscience and developmental brain genomics, clinical research, AI-modeling and data integration with youth engagement and ethical reflection. The authors of the current contribution are all members of this consortium. One of the central objectives of the consortium is to develop innovative models of prediction, prevention and interventions in youth mental health that build upon the above-mentioned complexities. To that end, it integrates the development of an AI-powered app for monitoring and self-detection to be used by both clinicians and patients. This technology is one promised deliverable of the consortium's research that will be tested and further refined in the upcoming clinical trial among help-seeking young people in 6 European countries (i.e., Spain, The Netherlands, Serbia, Croatia, Finland and Estonia).

In a unique collaboration of several members of the consortium, the aim of the current contribution is to provide insights into the (ongoing) development and complex decision-making that has shaped, and continues shaping, the generation and development of the AI-powered app. Research teams working on 'Engagement, ethics and networking' and on 'Data integration and prediction' as well as one of the consortium leaders and members of a dedicated youth advisory panel, aim to make visible and discuss the human decision-making process. We will describe who was involved in decision-making, which steps were taken in the preparatory, early and middle stages of technology development, but also which challenges were encountered, and the possible impact of these for the final product and the scope, and limits, of its usability. Together, we aim to increase understanding of human decision-making at the background of an upcoming AI-supported app in youth mental health care.

3. Approach

To that end, we invoke a multi-method approach, combining bi-disciplinary dialogues covering the ethicists involved and one further discipline respectively, with a transdisciplinary multilogue bringing all disciplines together and a subsequent material and document analysis.

3.1. Development and investigation of the Youth-GEMs-app

A mobile phone app is being developed to collect behavioral data from young participants in a clinical trial over a 24-months period. The app will gather digital markers about mobility, physical activity, phone usage, and sensor data on physiology (e.g., heart rate variability and sleep). Alongside the envisaged app-usage, clinicians will collect other relevant clinical and behavioral information on said participants by means of an agreed clinical protocol. The digital phenotype description collected through the app will serve as complementary data to the information collected by the clinicians. All the collected data (digital and clinical) will be used along with an ad hoc AI model to obtain mental health predictions. This approach aims to provide a comprehensive dataset, encompassing both traditional clinical data and real-time digital markers.

3.2. Bi-disciplinary dialogues with Youth-GEMs-app decision-makers

The two ethics members (DH & MK) in the research team took the initiative to investigate the decision-making underlying this app. To that end, they first identified those parties who were involved in the decision-making regarding a 'Youth-GEMs' app and set-up a series of dialogues or interview sessions in which the ethicists engaged in focused conversations with all these parties – the human decision-makers so to speak – involved. We investigated procedures in making decisions, described the various steps taken by participating parties and explored underlying motivations and reasons for concrete decisions. This includes two psychiatrists (BR & JvO) (BR who is currently one of two consortium leaders, both were involved in the very set-up of the research idea), a social scientist (SK), a group of biomedical engineers (MB, LG, MLG), a group of AI-researchers (NL, ERP, MC), as well as young advisors with lived experiences of mental ill health (IB, FR) and a representative (ES) from the organization of European Youth Mental Health. We held seven preparatory dialogue sessions in which the ethicists (DH & MK) consecutively spoke with the other disciplines individually or in small (uni-disciplinary) teams. Topics discussed covered all the respective areas of contribution of members in developing the app. In addition, the two psychiatrists and co-initiators of the research-project (BR & JvO) were invited to report about the background of the project and the motivation of the very inclusion of AI-tools; the more technically oriented members (MB, LG, MLG, NL, ERP, MC) reported about their interest in and view on mental health, and members who entered from a mental health perspective (SK, ES, BR, JvO) reflected on their view on digital and AI-supported approaches.

3.3. A transdisciplinary multilogue

An initial analysis for common themes and topics identified some diverging approaches and partly contrasting motivations and views. The common themes but also the partly diverging views formed the basis for one common so-called 'multilogue'. Multilogues are still a somewhat experimental approach to engage in and foster inter- and transdisciplinary collaboration and bring different views together. Multilogues are used to present and discuss one topic from the angle of different practical and theoretical approaches. In the context of mental health, especially regarding the recovery approach in the context of '*kwartiermaken*' (an initiative originating in The Netherlands that aims to stimulate participation of mental health clients in ordinary social life) (Kal, 2001), the term *multilogue* refers to a conversation where people from diverse backgrounds are invited to participate. The aim is to do justice to the very fact that mental suffering is intricately interwoven with everyday life and the social contexts in which people who suffer mentally live, being affected by their circumstances, but also exerting their own impact on their social environment. Therefore, for example "users of psychiatry, relatives, partners, and mental health care professionals, neighbors, professionals, trade unionists, human resource officers, politicians, architects, and policemen on the beat" (Mölders & INCA Projectbureau Amsterdam, 2013, p.1) come together to discuss cases. As equal partners in the discourse, they create a 'free-space', where a multitude of voices can be heard and everything which falls outside of 'the normal order of things' can be said safely (Kal, 2001). Multilogues have further been invoked as a means for narrative inquiry among participants holding diverse paradigmatic background assumptions to explain themselves and create a common ground of expression and understanding (Schubert, 2008). Meanwhile, they are a rather well-established means for developing and organizing transcultural competences in ethnically diverse communities to enable and facilitate cosmopolitan interchange (Hale, 2023). Our own international transdisciplinary consortium, with its different scientific paradigms, perspectives, habits and cultures of working resembles such transcultural environments and therefore can likewise profit from the multilogue approach. In this sense, our multilogue serves two aims. Regarding the procedure, it allows for a multivocal inquiry covering the different voices and perspectives active in the consortium; and in terms of content, it elucidates the characteristics of the various aspects and motives of human decision making on one common topic (the AI-supported mental health app under development).

During our multilogue, all participants from the initial dialogues met together and exchanged ideas. To foster this multilogual gathering, representative – sometimes contrasting/conflicting – quotes from the initial dialogues were selected and presented to everybody. Quotes were selected on the basis of emerging three main themes and quotes that were more likely to elicit debate in the multilogue were preferred above

those that appeared more agreeable among the contributors. This allowed the different team members to learn about their mutual views and contributions, and to react upon each other in equity. While the ethics-members of the consortium DH & MK initiated and organized the various interchanges, they aimed to stay on an equal footing with their colleagues in the consortium as far as possible. They had a central role during the dialogues and in the initial analysis but stepped back and merely moderated during the multilogue. The contributing parties hence are not participants in a traditional focus-group or interview setting, but co-researchers and co-authors of the current contribution. For this reason, no ethical approval was needed and informed consent procedures are not applicable. All dialogues and the multilogue were held online, making use of the online video-meeting platform Zoom. Meetings were recorded and recordings were shared with the respective participants.

3.4. Material and document analysis

In addition to, and based on insights from, the verbal communication, we also revisited the work-in-progress material presented and discussed during several co-creation meetings held in the context of the app-development among the AI-researchers, the biomedical engineers building the app and particularly the young experts by experience.

The two psychiatrists, who were also involved in the initial phases of the current research project BR and JvO, indicated immediately that making decisions on the inclusion of AI-modeling and data processing did not start with them. It was inspired by current developments in contemporary data analysis. In addition, it was called for and urged by the research call set-out by the EU Horizon program, which in turn was based on EU-policy decision-making procedures. For funding, applications must closely live up to the set requirements. In this sense, the call itself and its emphasis on the inclusion of AI-tools should be considered a preceding moment in the chain of human decision-makings involved in the AI-supported app under investigation. To do justice to this aspect in the very early human decision-making process, we conducted document analysis of the underlying call text, the original grant-proposal and the grant reviewer's comments.

Against this background, this publication will provide insights into backgrounds and the course of the various steps of human decision making in the context of the AI-supported app under development in our common consortium. To explain and justify any steps and decisions taken, we will also explore our motivations to engage in AI-development, our respective expectations, views, and hopes, but also critical considerations regarding underlying assumptions or potential structural limitations of the very integration of AI-supported mechanisms. Given the diversity of our own group, we do not aim for full agreement, or a coherent view on the – potential – role of AI in youth mental health care, but we do aim for transparency, intelligibility and open mindedness.

4. Human decision-making in the forefront of AI-research

AI-modeling and AI-supported decision-making are gaining increasing attention in medical and mental health research. Even though most research on AI-supported decision making is currently taking place in the field of radiology and medical imaging (Goisauf & Cano-Abadia, 2022; Hosny, Parmar, Quackenbush, Schwartz & Aerts, 2018), other domains of health care, including mental health care, aim to make likewise use of these innovative forms to analyze huge amounts of data and develop new predictive models and means of personalized medicine. This trend is driven by a genuine scientific interest and the hope and expectation that more reliable findings and personalized predictions will become possible and it is reflected by the scientific research policy of the European Commission (Correia & Reyes, 2020) that aims to advance this 'game-changing technology' (p.3) and boost the 'new electricity' (p.5) of our times.

Against this overall background, the EC commissioned the Horizon call 'Staying Healthy 2021' (European Commission, 2021) that finances the current consortium. This call requested grant-applications to deliver, among others, the following specific outcomes: better understanding of how genetic, epigenetic and environmental risk and resilience factors *interact* to drive or prevent the transition of mental health to mental illness, and optimizing and *personalizing* preventive and therapeutic treatment decisions. This request to identify complex interactions and use this knowledge for personalized decision-making, departs from the assumption that 'a deeper molecular and neurobiological understanding of the interplay […] is critical for the development of objective biomarkers and evidence-based interventions" (European Commission, 2021, Scope). This focus on the molecular and neurobiological dimension of mental health together with the expectation that multiple interactions between genetic, epigenetic and environmental features are going to deliver enormous amounts of complex data in turn motivates the explicit encouragement to make "use of computational modeling and/or artificial intelligence tools" to achieve the purpose of the "development of robust *quantitative* [emphasis DH], clinical measures in mental health" (European Commission, 2021, Expected Deliverables)..

The current consortium is based upon a successful research application to the Horizon Europe funding scheme. Responding to the call-requirements, the initiators of the proposal decided to engage in the federated analysis of existing cohorts for the development of data-driven prediction tools and the application of machine learning and AI to develop AI-driven prediction tools for young people and clinicians (cf. Table 1).

These elements in turn received particular positive attention by the – anonymous but still human – reviewers. Scoring the criterion of excellence, the review report emphasizes that the proposal performs 'machine learning and artificial intelligence (AI) analyzes […] on determinants and phenotypes of mental health trajectories in >40.000 young people', states that 'it combines state-of-the-art approaches including AI' and concludes by highlighting that 'the quality and robustness of artificial intelligence are well addressed using validated computational models for complex data integration'. Even though their contribution is at a distance and reviewers' task is to assess applications to the call text, their evaluation is also a crucial element in the human decision-making that underlies the current development of the AI-supported app for youth mental health care.

Table 1
Expectations by the funding call (European Commission, 2021) and reply by the research proposal (excerpt): [left column, rows 2 and 3 quote from funding call, recapped in research proposal; right column, rows 2 and 3 quote from research proposal].

What the call expects research projects to deliver l	How the scope of the research proposal aims to meet the expectation
"Develop relevant predictive models through federated analysis of large European cohorts of psychiatric disorders and investigate the biological and neural basis of pathogenetic mechanisms and symptoms"	• Federated analysis of existing large European longitudinal youth cohorts and development of predictive models (identifying actionable targets) for more precise trajectories of mental health and illness beyond current symptom-based criteria • Development of data-driven prediction tool for help-seeking young people and clinicians.
"The use of computational modeling and/or artificial intelligence tools is encouraged for the analysis of big, complex and heterogeneous data"	• Apply machine learning and AI to: ○ Identify trajectories of mental health in young people. ○ Identify sets of (early) markers of illness throughout the critical age range. • Develop AI-driven prediction tools for young people and clinicians, and test these *in silico*.

In sum, the call text unequivocally requested inclusion of AI-innovations, the grant-proposal clearly responded to this request and the reviewers emphasized the value of the proposed quality of AI. Still, among the project's initiators BR and JvO the motivations why and how to include AI-tools appear diverse. Apparently, from the early decisions taken on AI-usage, their different motivations reflect the call's aim to combine a biomedical perspective on mental health with a humanities and experience-based view. On the one hand, psychiatrist BR expected that making use of machine learning to analyze the great amount of already available data will have an added value compared to traditional statistical means of analyzing, may possibly identify certain (yet unidentified) factors as candidates for possible predictors of mental health status, and thereby potentially be able to trace back causal pathways of known risk-factors. In addition, the AI-tool included into the app and used in the consortium, might then validate, or maybe refute, the meaningfulness of these newly found variables in the planned clinical study of the consortium. On the other hand, psychiatrist JvO focused mainly on young people themselves and their experience of mental suffering. The initial idea to rely on an AI-element in an app for young people at risk or seeking help for mental distress, may not be helpful in itself. Instead, young people should learn to reflect upon themselves and their emotions. However, a technical AI-tool cannot show them something about their inner selves, so to speak. Young people themselves are the intelligence-force, so to speak, that should be at work in mental health care. In this contested context, it might be particularly worthwhile to have AI-developers and young experiential experts work together, because that allows these different perspectives to learn from each other. Technically oriented AI-experts can learn first-hand what it means for young people to suffer mentally, and young people with expertise by experience can impact ongoing real-world developments in mental health interventions. Deciding in favor of the inclusion of young people's experience, as well as a social science and humanities perspective introduces a dialectic perspective, allows for mutual learning processes and can do justice to the current epistemic pluralism on mental health.

5. Transdisciplinary multilogue on human decision-making in AI-development

These decisions did not only substantiate the very decision-making on an AI-supported mental health app, but also shaped our common work in the multilogue on the three themes of: 1) using AI, 2) co-creation in app development and 3) the app, autonomy, and human relations.

5.1. Using AI

During the initial dialogues with consortium members, we talked about their reasons and motivations to make use of an AI approach at all. Why are AI-tools included in the work, what does one expect from AI-modeling in general and in addition to traditional means of investigation and analysis? But also, how does one make use of AI in the current setting and is this similar or different to the usage of AI in other settings, if applicable. The following quotes from these initial dialogues introduced this theme to the subsequent multilogue.

JvO, psychiatrist: "The EU wants to stimulate cooperation between different researchers from different countries and cultures. But biomedical researchers have no idea of the impact of what they are doing with all this data. It's important that they can see that this is not about data but about real people who struggle with issues".

NL, AI-expert: "Psychiatrists told us: this is what we really want, we want to know how well this person will be in the future. With AI we can detect those participants who have troubles in the future, we can choose and select those people and monitor them and the psychiatrist can work with them to improve their mental health".

BR, psychiatrist: "On the basis of big data sets and analysis of these data we can develop predictive models that work better than a hypothesis-oriented approach".

FR, initials young expert by experience: "It's better to learn how to seek human help and build a support system than relying on AI. I guess going forward the world is going to be way more digitalized, but for me in terms of mental health and preventing mental illness, it's really important to have built up a [human] support system around you".

Consortium members expressed different views on the reasons for and the potential goal of making use of artificial intelligence for youth mental health care. During the multilogue further explanations and explorations on this theme were given. Consortium leader BR emphasized that there are many data sets available, but no one had ever tried to use machine-learning to figure out whether that might help to formulate better predictions of trajectories of mental health in young people. Even if there are limitations and challenges; not trying at all and not making as good use as possible of the data already existing is also problematic and might result in blameworthy negligence. AI-expert NL further explored that the close collaboration between the AI-team and psychiatrists has been particularly essential and from their perspective far more in the current mental health project than in previous somatic medical-AI projects. To even be able to do their work and engage in modeling, AI experts need the input of psychiatrists to explain the many features and variables that affect mental health conditions and to support the *multi*-modeling process. At the same time, upcoming AI-based evidence hopefully can support psychiatrists to have a prognosis for their patients. From an AI-perspective, work in the context of mental health essentially appears to be of greater complexity, requesting more close collaboration with medical experts and designing more complex multi-modeling to overcome the current situation in which models are not yet good enough, and also not yet validated in different populations *(MC)*. However, if these struggles were solved in some future, AI-models might still become a trustworthy source of information for psychiatrists and be used as an advanced assistant tool for prediction and monitoring by psychiatrists and other mental health specialists *(ERP)*. But, as social scientist SK remarks, making predictions in mental health care is in a way a kind of reductionism. AI is human made and it builds further upon human decisions. Although, or because, it can evolve, we must be aware of this. In this context, fairness and the unbiased inclusion of the data of, for example culturally- and/or gender-diverse populations are considered crucial in our consortium, as it had been in the current AI-ethics literature (Cary et al., 2023; Hafner, Peifer & Hafner, 2024; O'Connor & Liu, 2024; United Nations Educational & Organization, 2020) and the EU-AI-policy requesting attention for inbuilt bias. However, even if this issue was solved and an algorithm that covered all cultural, gender, and further relevant dimensions was developed; the question remains open what this means for the prediction of the young people's personal experience, or as JvO had put it 'about real people who struggle'. Insofar as personal experiences are non-quantifiable, they are beyond the reach of even the most complex form of datafication. This reduces algorithms inevitably towards the quantifiable and it is at least an open question whether this entails a potentially unavoidable bias of AI in mental health.

Against this background, our multilogue on the use of AI does not so much indicate disagreement on the value and possibilities of AI in ethical mental health care, but rather sheds light on our different philosophical background assumptions on what makes up mental health and mental illness. As far as mental health is considered essentially due to molecular and neurobiological disturbances, AI-modeling can potentially be a support tool like it is to molecular somatic conditions. However, when mental health is mainly conceptualized as a mental phenomenon of experience, thoughts, and feelings, AI-modeling gets challenged on conceptual grounds. As JvO, following Foucault's understanding of medicine (Foucault, 1973), explored during the multilogue, the concept of 'epistemic pluralism' renders this diversity in perspectives intelligible. The psychiatric gaze (based on Foucault's description on how doctors filter patients' stories through a biomedical paradigm) in which biomedical features, interventions and risk- and

protective factors are paramount are one way to look at mental distress and gain knowledge (episteme) on mental health. This psychiatric gaze covers quantifiable features and allows for, albeit highly complex, AI-modeling. However, it is different from the perspective of experience and the perspective of the person who experiences problems. This other gaze covers non-quantifiable features and cannot be caught in AI-modeling. Given the diversity of the backgrounds of the consortium members, at this point, we will not further discuss the respective value or plausibility of the one or the other gaze. Instead, we limit ourselves to the observation that these two gazes exist and have a fundamental impact on our decision whether, and if so how, to decide on the use of AI-tools in mental health prediction, monitoring and intervention.

5.2. Co-creation

After having discussed the background of making use of AI, underlying motives and reasons as well as actual procedures, we focused on the decisions taken by consortium members in building the app and its AI-element. This is largely done in the context of co-creation, where app-developers, AI-experts, psychiatrists and young experts by experience who provide advise collaborate and together aim to shape the tool and fine-tune it. The following quotes from this subgroup of collaborators set off the common reflection on co-creation.

MLG, biomedical engineer: "Our aim was to involve young advisors in all the phases, initial designing, share and ask for feedback, all aspects of the app, start with very generic questions and then more and more specific, to make sure everything was taken into account".

SK, social scientist: "The young experts who cooperate in this project as advisors are not representative for all future users. They mostly are highly educated and interested in participating in research, but even then, I wonder if they understand everything about how the app is meant to work and for what purposes".

ES, representative of young advisor organization: "Young people with diverse ethnic and cultural backgrounds experience more barriers to access and so are less present in mental health care".

FR, young expert by experience and advisor: "There wasn't much time to talk about if there is attention for neurodiversity, for example, I was thinking about the issues of recognizing voice patterns, some people are neurodiverse, people with autism can have different patterns in voice, my concern is how would that be taken into account".

IB, younf expert by experience and advisor: "We were given access to a document where we could give feedback on different parts of the app like the graph, the content, the words, the colors, the images, what to change and what was good".

MLG, biomedical engineer: "Some things were really highlighted by the young advisors as very fundamental and at the beginning we could not think about these things as the most relevant issues".

Co-creation is an important element in human decision making in research projects, pointing towards the ideal that not only scientists are taking part in research decisions, but also potential users or their representatives. A multidisciplinary team of scientists conceived the Youth-GEMs project. Young people with experiential expertise were invited to the project at its very start, but after the conceptual phase. This timing might be an inherent limitation (JvO) where young advisors may only advise on what – in general lines – others [researchers who write grant proposals] had invented.

At the same time, co-creation requires integration of young people from the very development of an initial idea and collaborative work throughout. One challenge is ensuring everyone has been trained to have the relevant skills or knowledge, so in this sense consultation can sometimes be easier and immediately experienced as meaningful by the co-creators (ES). While agreeing with the limitations of the current approach, also the app-developers pointed towards the inherent challenges of having co-creation at even earlier phases. "Maybe we have to rethink co-creation, because now we use it in normal development practices where both parties (developers and users) share the starting point. But in research, this is different. Here the young advisors did not contribute to the starting point. They were not the ones asking for an AI-powered app. This, however, is quite typically in early research stages where "someone" has an idea. Research includes discovery phases. To me it is hard to understand how scientific discovery may be co-created (even if it might be helpful)" (MB, biomedical engineer). Supporting this point, the young people declared themselves to be well satisfied with their current involvement in co-creation. "It's really good that we've had involvement and we're getting feedback on it" (FR) and they declared being highly interested in the AI-aspect and curious what it might deliver. Against that background, it remains unclear whether even earlier common decision-making, i.e. including also co-creation during the starting phase of the project and on the very question of developing an AI-supported app would have been preferable. It also remains unclear what such a step might have entailed, and whether and how it could have been meaningful for both scientific and lay co-creators.

Taking these considerations on the timing of the start of shared human decision-making into account, this part of our multilogue covered two main topics: i) who takes part as a co-creator and ii) which elements of the AI-supported app are co-created?

5.2.1. The role of young experts

As a background, it is essential to acknowledge that in the current consortium, the group of young experts with lived experiences was included in all research teams and all aspects of the work done, of which the building of the clinical app and its AI-element was just one part. In this sense, the overall co-creation process can be considered rather broad, meaning that decisions in all parts of the research performed by the consortium can take the view of young people into account. To this aim, there is a monthly online so-called 'Science Café' where researchers share their activities and questions with the group of young advisors. In addition, young advisors attend the regular (online) meetings of the research teams to keep being informed (as far as possible). This broadness of influencing decisions, however, comes with the challenge of enabling meaningful participation and co-creation. Social scientist SK, and the young advisors FR and IB themselves easily agreed that the young people who joined the consortium represent young people with mental health experiences only to a limited extent. They are all young adults rather than children or teenagers, currently all are involved in higher education, fluent in speaking English and interested in thinking about mental health issues. While this might be a limitation and lack of overall representativeness of the group of mentally distressed young people, there are also affirmative reasons to make decisions together with exactly this group. Biomedical engineers MB and MLG emphasized that the above-mentioned characteristics of the young advisors are important to enable meaningful collaboration and co-creation in this early phase of doing research. Because of the existing contacts with Euro Youth Mental Health (EYMH), a group of young advisors was available, enthusiastic and eager to participate, and actual collaboration was feasible. Moreover, once the consortium was in place, new young people were recruited with the project's specifications and requirements in mind. Rather than explicitly excluding any young people, a decision was taken to start actually consulting young advisors, to give them a voice and options for input and to allow the researchers and app-designers to tailor the app according to these advisors' input. "We needed to develop something that may be used by younger people [..] and that is trusted and accepted by them" (MB). The current group of young advisors, even though not representative in all regards, allowed for this immediate and effective consultation. MLG provided a concrete example of this co-creation and how it helped the engineering process: "When we started designing the app, we had some ideas of features that could be included and we immediately asked young advisors for their feedback and also to provide ideas completely from scratch without having any bias from our previous thoughts". In this sense, essential decisions on app development were taken, and changed, based on the advice of young people. For example, the app intends to entail a function where a daily question

prompts users to report about themselves. These questions were extensively commented on and often changed as a consequence. For example, the original question "*If a genius would grant you three wishes, what would you like*?" was rejected because when thinking about it, a user – who all will be young people in mental distress – might focus on what is lacking in their life, being a trigger for those young people who already hold negative thoughts. The question "*What do you admire in another person*?" was welcomed as it echoes the idea of being inspired by other people, rather than negatively comparing oneself to others. On this concrete level of the development of the app-interface, i.e. the questions users will be asked, the colors they will see, and the experience they will undergo using it, the common decision making was extensive and biomedical engineers and young advisors operated iteratively and on an equal basis.

But on the level of mathematics and coding, there was no co-creation. As the AI-experts explored, coding is something that asks for specific expertise and knowledge. First, there has to be some baseline model (MC); discussions on fairness, bias, etc. come later. In this later design and development phase co-creating is feasible and highly valuable to avoid developing something useless..

5.2.2. The role of multistakeholder groups

This, however, leaves open the question of who takes decisions on the actual AI-element of the app under development. This interface, which should be attractive and inviting to young people to foster interactions, is only one part of the app. The other part consists of an actual AI-element. This element is currently under investigation and covers the analysis of lifestyle and behavioral data as well as voice and typing data gathered during app usage to investigate whether these data are indeed useful for monitoring and could be used in the future as an element in the prediction of mental health. Analysis of the data is done making use of traditional supervised machine learning techniques (e.g., Balanced Random Forest, XGBoost, Support Vector Machine, and Logistic Regression) where algorithms are trained on labeled data and with trustworthy metrics including explainability and fairness metrics, or unsupervised clustering techniques to identify patterns (e.g., k-means for time series or latent class analysis).

In the actual AI-driven element, co-creation took place with health care professionals including psychiatrists who had been extensively consulted on their needs for a model. Co-creation did not take place with young advisors and there is general agreement that this is not possible. "Without a background in AI, I can't begin to even imagine where we (EYMH) would slot into those places" (ES). In a similar vein, biomedical engineer (MB) explored that and why co-creation in this part of the work might not be easily possible, even with a rather well-educated group of young advisors. "It's a different step, there are papers that say that there is scientific evidence that points out that some behavioral data are meaningful, but scientific evidence is relatively small, and we first have to research. Therefore, this cannot really be co-created" (MB). AI-developers agree that during the early phases of research in AI-modeling, co-creation with young advisors is not feasible. "We are still not in the moment of adding them [young advisors] in the co-creation right now, because there's a lot of coding that has to happen before asking for advice" (MC). So, while there was quite some decision space for young advisors in the development and design phase, at the very early phase in the research on the AI-model itself, there appeared to be no meaningful room for the input of young advisors.

However, as discussed during the preparatory dialogues particularly with the AI-experts NL, ERP and MC, a multi-stakeholder co-creation group was supporting the work and particularly the psychiatrists in the consortium participated in the more low-level decision-making. The psychiatrists' professional expertise seemed even more important in AI-modeling for mental health than the expertise of somatic medical doctors were needed for AI-modeling in previous projects, for example on cardiology, carried out by the same AI-experts. Psychiatrists were essential at all stages of the AI modeling process, from discussing the study's primary objectives and determining the most relevant variables, to understanding the model's decisions and providing clinical interpretations. In terms of study definition, they emphasized the importance of not focusing solely on a single mental health disorder, as mental health conditions often coexist and share underlying factors, an exploration that exhibits the consortium's innovative trans-syndromal approach to mental health (Fusar-Poli et al., 2019). AI-developers spend much time to understand mental health trajectories (e.g. by generating statistical models of the trajectories) and what defines mental health for patients and to that end relied extensively on the input of psychiatrists to understand the relationship between specific data points and mental health outcomes. At this stage, young advisors, however, did not engage in the discussions. Their view was represented via health care professionals; however, it is unclear whether their direct input would have led to any changes. In the final stages, psychiatrists, but not young advisors, also played a crucial role in interpreting how the features affect the model's outcomes, helping to validate clinical relevance and support the model's applicability in real-world mental health contexts. Psychiatrists' expertise advises AI-driven insights to align with clinical observations and inform therapeutic strategies and patient care. In sum, the process of modeling, validating, and re-modeling is conducted in close collaboration between AI-developers and psychiatrists and decisions are steered by available data sets as well as the requests and explanations provided by psychiatrists.

5.3. App, autonomy and relationships

The AI-ethics discourse pays quite some attention to ethical principles and values and how AI-tools might be invoked to undermine or support such principles. Debates on ethical compliance, in turn, mainly focus on the question of how AI-tools are to be designed to ensure that they live up to ethical requirements. Even though our multilogue is broader and more bottom-up, the principle of autonomy, the corresponding ideal of empowerment or self-empowerment and the role of fostering relationships with other humans, was emphatically and controversially discussed during the dialogues. In the multilogue, we reflected back on the different positions on these themes held by the respective contributors in the multilogue.

NL & MC, *AI-experts: "The project is controlled with privacy agreements and ethical approvals, the model does not use personal data but pseudonymized data, … users can choose which data they want to share and which not.*"

ES, Representative of EYMH: "In the Science Café, o*ne young advisor came up with the suggestion to maybe not explain that we are looking at your typing pattern or your voice [at your behavioural and mobility patterns] because people will be too consciously aware and might change those things being aware that they are being watched in a way.*"

IB, Young expert by experience and advisor: "*Some people don't have the possibility to talk about their mental health, so I think that having an app that can help you to keep track of your emotions and well-being, it can make you feel more validated."*

SK, Social Scientist: "*Anxiety can get worse when there is a constant focus on the phone; my fear is that people with anxiety problems get more anxious when they constantly have to be alert and get primed by push-messages and questions*."

MB, Biomedical engineer: "*The next step is to have this information not only as something that the patient collects and sends to the clinician but to make information that may be used to empower users, to go to self-management. […] The app is becoming a helper to give me recommendations, or to give me some information, so we move to self-management. It's an evolution in technology, it's fine.*"

During the dialogues, we discussed extensively the potential impact that the usage of the AI-supported app currently developed might have for future users and how clinicians should instruct young people. The background of this theme lies in the extensive ethical debate on the potential impact on autonomy and empowerment that mental health

apps in general and those supported with an AI-model might have on young people in mental health care. Some hope that mental health apps are more easily accessible and available whenever needed, improving autonomy and enabling empowerment (Martinez-Martin et al., 2018), whereas others are concerned that such digital tools might exert pressure on users, further increase already troublesome phone-usage , and disturb the therapeutic alliance that seems particularly essential for mental health care (Desmurget, 2022; Haidt, 2024; Bedor Hiland, 2021), ultimately undermining user's autonomy. This general debate is also reflected among our group, where concerns about increased phone use (SK) are accompanied by the trust that a mental health app might be a particularly accessible form of support (IB), and a tool that can install or support self-management because the treating clinician is better informed about the person's situation (MB). An interesting difference in view seems to exist between the AI-experts and the youth advisory panel. In line with current AI-regulations and AI ethical principles (Jobin, Ienca & Vayena, 2019), AI-experts seek to ensure that the app is compliant with ethical and judicial requirements of informed consent, data-protection and the overall GDPR guidelines and hence allows users to choose which data to gather and share (NL & MC). Notably, ES reported that members of the panel of young people advised to maybe *not* inform users on everything, because otherwise one might get overly conscious about those data gathered, get anxious, react differently than if one was naïve, and thereby inadvertently undermine the meaningfulness of the data gathered.

This debate on the impact of the Youth-GEMs-app on autonomy and empowerment of its users continued during the multilogue and no consensus was achieved. SK explored further that the app might introduce additional stress, induce a fear-of-missing-out on activities or constantly remind people of their unbalances, drawing their attention to their – troublesome – mental states. FR reported from the first testing-phase of the app that there was indeed a pressure felt to use it and to regularly carry out the various activities requested by the app. The feature that requests a user to speak aloud even requires them to look for a quiet private space, which can make usage difficult. For the developing team such experiences constitute a reason to ensure that there is no requirement to do the tasks every day and users are free to choose which activities to perform (LG).

From various standpoints, MB's initial quote that the app should empower users, support self-management and become a helper, is broadly supported during the multilogue. Still there are different views on which decisions are needed to potentially achieve autonomy and empowerment. JvO pleaded for an AI-free app that mainly reports back to users on, for example, their emotional fluctuations, so they can use this information as an invitation to reflect on oneself. This enacts a view on autonomy that mainly focuses on the reflection and agency of the individual themselves. BR rather expects that additional AI-modeling might be a tool for the health care professional (e.g. a psychiatrist) to be better informed about the help-seeking individual/patient, have improved information on patterns of experiences and context of the patient. This should enable better discussion and reflection together with the patient based on their own (passive and active) input, and professionals may therefore be better equipped to support patient's empowerment. While this attitude likewise makes reference to terms of autonomy and empowerment, it rather views empowerment as a condition that requires the active involvement of another person who in turn depends on information provided about the individual, rather than by the same individual themselves. Essentially this broad agreement on the value of autonomy, empowerment and good care, brought the discussion back to where it started, i.e. the (limited) value that AI-modeling in presenting findings about themselves can have for the young people who are using the app. In the end, however, this again comes down to the question on the concept of mental health, and the role of personal experience in it, rather than a question of which exact AI-model is used to process the inserted data and come to a personalized prediction. A concept of mental health that focuses on personal experience and attitude towards one's own emotions, behaviors and way of being, is not in need of a complex algorithmic calculation of a plethora of behavioral, emotional, physical and social features; while a concept of mental health that aims to make sense out of a broad spectrum of mental health symptoms, will see potential benefit of such calculations, and even more so when supported by encompassing AI-tools.

6. Discussion and conclusion

Our analysis of the research call and grant-proposals as documents that constituted the current research, as well as our current series of bi- and transdisciplinary conversations pointed towards a complex decision-making process in the development of an AI-supported mental health app. A variety of actors are involved, and different spheres of decision-making have become visible. In the following, we will reflect on these insights.

6.1. Strengths and limitations of the transdisciplinary multilogue approach

In the current study, we aimed to make visible the human decision-making behind an AI-supported app developed in the context of a European Consortium on youth mental health care. Given the different disciplines involved in the consortium and the process of app-development, we investigated the various decisions taken by the different actors, the interrelationship between these decisions and between those who make them, and the respective scope and limits of everybody's contribution. This allowed us to not only render human decision-making behind AI more visible, but at the same time, the multilogue turned out to be a didactic play on the possibilities but also the challenges and limitations of transdisciplinary work with a diverse team of researchers covering ethicists, social scientist, experts in AI and psychiatry, industry leaders and young people with expertise by experience. Even though, we explicitly aimed at equity of all participants, we should not ignore, or consider ourselves immune to, still existing hierarchies as regards seniority of everybody's contribution, but also different levels of the respective epistemic authority of all participating disciplines and stakeholders. We explicitly encountered different understandings of the scope and meaning of mental health issues, with some of us focusing mainly on the experiential dimension, while others preferred to emphasize biomedical features as those that ideally can be modelled for algorithmic purposes. This very debate, of course, is not new. The meaning and bearing of 'mental health' has been the subject of many debates (Singh & Rose 2009). Biomedical (and as a consequence 'model-able') approaches have been suggested by prominent experts in psychiatry (Klengel et al., Musci et al.), as have been rather social science and humanities informed understandings of mental suffering by others (Aho, 2008). Oftentimes, these more broad debates are characterized by a significant epistemic power imbalance, where biomedical and technological approaches are given more credit and are perceived as more objective because of their quantitative, data-driven and technological character. Experiential and social science approaches instead may be considered more subjective and interpretative and ascribed less credibility. Despite the clear intention of each contributor to the multilogue to strive for equality and inclusion we cannot rule out that processes of implicit bias that reflect established scientific hierarchies also shaped our multidisciplinary work and made it more difficult to speak up, or to challenge dominant narratives, for those who represent less dominant voices.

At the same time, we acknowledge that we focus on the work of just one specific consortium and the development of one specific AI-supported mental health app. At the current moment, it is unclear whether, and to what extent, our thoughts and insights are equally valid for and generalizable towards other AI-tools in other mental health research projects or in the context of somatic healthcare. Moreover, the agenda of the multilogue as well as the preceding dialogues had been set

by ethicist contributors, and in that sense, it was not co-created equally by all partners. To ensure focus on one concrete aspect of the work performed by the overall consortium, we involved only two, out of totally seven, of the consortium's research teams. However, further and potentially different insights might have been gained if representatives from all research teams had their input and if further AI-related activities, beyond the development of the Yoth-GEMs-app, had been analyzed. We finally emphasized the views and perspectives of team members on the development of AI and analyzed the development and design of the app to some extent, but did not dig into the many technical decisions. This might limit our analysis; at the same time this selection also allowed us to focus on the broadness of human decision-making behind AI.

Based on the initial dialogues with the project's initiators, we iteratively decided to also include a document analysis of the research proposal, its reviews and the funding call text. However, we did not engage in interviews or dialogues with the creators of the funding call and hence have no first-hand insights into the motivations of the designers of the call from the European Commission to even request the inclusion of AI-development in youth mental health innovation.

6.2. *Distant decision making: the role of EU research policy requirements and regulatory frameworks*

During the preparatory dialogues with the initiators of the research project it became obvious that decisions regarding the AI-supported youth mental health app are being taken not only regarding *how to* design, develop and integrate the AI-tool, but more fundamentally that preceding decisions had been taken regarding *that* an AI-supported app should even be developed and integrated in innovative youth mental health care. In this sense, the role of the funding-call text as designed and decided upon by the European Commission must not be underestimated. As a consequence, researchers who aimed to engage in innovative youth mental health research did not have the decision space to decide about the very development and inclusion of some AI-supported tool. Obviously, several members of the consortium and contributors to the multilogue are very much in agreement with this requirement and see much potential in the development of an AI-supported app and its clinical investigation for better youth mental health care, for the prediction of upcoming mental health trajectories and for the specification (and ultimate tackling) of known general risk-factors. While the EU research policy, exemplified by the current research call, enables human decision-making on the development and design of the Youth-GEMs-app, it discourages human decision-making on tools which emphasize and support the experiential dimension of mental ill health and distress. Such tools might make use of analogue or digital means, but cannot depend on algorithmic insights, as these require quantitative data.

But also, the inclusion of quantitative mental health data into an AI-supported app is a challenging endeavor as our research teams experience. When translating mental health expertise into algorithmic decision making, valid, reproducible and meaningful analyzes and predictions must result from this process. This process and all decisions taken along the route are at least influenced, maybe challenged, but perhaps also structured and supported, by set regulatory frameworks that steer the collection and processing of any underlying data (e.g. data-sharing and data access regulations). These legal regulatory and judicial frameworks, such as the EU General Data Protection Regulation (Wolters, 2018), or the Helsinki Declaration (WMA, 2024) on scientific research with human subjects, were not at the heart of any of our conversations, nor were they explicitly mentioned by any partner. Probably because it's obvious that, in the background of development, regulations on data-protection and informed consent in research and treatment settings must be respected and technology developers and researchers can only take decisions that are in compliance with these regulations, but not critically discuss these regulations as such.

The requirement of 'informed consent', for example, entails that users of the app know and consent to the kind of data gathered, the purposes of the data gathering and the way their personal data will get processed. At the same time, not informing users or not providing them with the option to select which data to gather and share would also break users 'rights as defined by the European General Data Protection Regulation (GDPR). However, at some point, young advisors in the consortium argued that it might be preferable to NOT explain to users all the ins and outs of passive and active data gathering, as this might steer and actually change these users' behavior, resulting in less reliable data on, for example, their typical physical activities, the way they answer questions asked by the app, or how they engage socially. Less reliable data, in turn, run the danger of resulting in less helpful outcomes and thereby challenge, if not undermine, the main goal of the app, that is, to restore and increase youth mental well-being. On other occasions, young people raised concerns about over-surveillance by an app that passively monitors many of their whereabouts, and also about stress and anxiety triggered by the very requirement to actively enter data into the app. Importantly, young people's advice and the options of developers to design 'freely' are confined by set regulatory frameworks.

It, however, goes beyond the scope and the aim of the current article to further discuss the justifiability and limitations of these regulations in the context of youth mental health innovations, but it should be acknowledged that in all current decisions, such judicial requirements and the ethical dilemmas they might evoke, must be taken into consideration. From a judicial point of view 'compliance' is crucial and unavoidable, however from an ethical point of view, issues are not necessarily set by being compliant to guidelines, and new dilemmas and questions can result (Bak, Ploem, Tan, Blom & Willems, 2023; Marelli, Lievevrouw & Van Hoyweghen, 2020).

6.3. *Proximal decision-making: across different disciplines and together with stakeholders*

The broadness and transdisciplinarity of the research team making up the consortium is the first clear indication of the variety of actors involved in human decision-making on AI. What's more, the technological experts in the team emphasized how crucial their exchange with mental health experts and psychiatrists was for their analysis of the sourced mental health databanks and hence the modeling of the data. In this sense, clinicians argued that mental health assessments such as the Strength and Difficulties Questionnaire (Goodman & Goodman 2009), life tracker or measures of social functioning were more meaningful than focusing on single diseases. These assessments are designed to evaluate a range of symptoms that may overlap across conditions, providing a transdiagnostic approach that better reflects the complexity of mental health. Thereby, psychiatrists and other health care professionals turned out to be an essential group of decision-makers steering the development of the AI-tool. In the current context, this is particularly crucial to acknowledge, as the understanding of mental health is characterized by a significant conceptual shift, abandoning the traditional focus on differentiated disorders and emphasizing a transdiagnostic approach of mental distress. It is this latter approach to mental health that also steers the technological decision-making in the mental health app. Therefore, the collaboration across the disciplines of psychiatry, bioengineering and AI-modeling can be an instance of full-fledged co-creation (Vargas, Whelan, Brimblecombe & Allender et al. 2022). Without the input of the psychiatrists and their transdiagnostic conceptualization of mental health, the AI-supported Youth-GEMs-app would look and function very differently, and without the concerted efforts of bioengineers and AI-experts, psychiatrists would be devoid of the potentials of algorithmic insights in their attempts to approach young people with mental distress transdiagnostically.

Decisions were also taken together with young experts by experience in order to ensure meaningfulness and attractiveness of the app from a user's perspective. All parties agreed that the involvement of young experts improved the development and design process, had a clear

impact on how the final app is presented to later users, and has the potential to overcome known barriers of implementation of digital mental health in real-world settings (Graham, Lattie & Mohr, 2019). However, we still encountered challenges in balancing co-creation with young people and the technical complexity of AI development. While co-creation with young experts was extensive during the development and design of the interface, the interactive modules and the way the app appeared to its users, co-creation did not reach the level of modeling or coding, and it did not cover the very question of whether an AI-supported app should be developed in the first place. Actually, the kind and level of participation of particularly the young experts by experience was rather diverse. At some points, they took part as equal participants in the process or even determined certain steps in app-development; at other moments, however, they were kept outside decision-making and were merely informed or consulted on their opinion. Despite intensive efforts to include all stakeholders and disciplines, technical and language barriers persisted and not everybody was equally able to use or understand technical terminology, complex statistical models, or the early steps in coding and modeling, limiting the ability of some, mainly young advisors but also social scientists and ethicists, to meaningfully engage in these parts of decision-making. As a consequence, there were different levels of participation by particularly the young advisors and non-technical contributors to the consortium. Theoretical approaches to participation typically make a sharp difference between, for example, consultation and co-creation, or partnership (e.g. Arnstein 1969, 2019) for the analysis of the role of stakeholders in any design or policy process. In our multilogue, however, we identified that the same stakeholders, young experts by experience and advisors, are partly involved as full-fledged equal, or even controlling, co-creators and partly they have been consulted or merely informed about decisions taken. Therefore, our multilogue points out that and how lower levels of participation, such as information and consultation, can co-exist with more full-fledged partnership and co-creational control. This is further indicated by a great degree of satisfaction among the young advisors with their role and the way they feel seen and heard when, most notably, developing the interface, appearance and modes of interaction of the app.

Acknowledging these positive experiences, we further discussed whether, and if so, how it might be feasible to also engage in full-fledged co-creation, rather than mere information or consultation, when exploring something that is totally new for everybody, and in the earliest phases of coding and design. Previous studies likewise indicated structural challenges on the involvement of service users in the design phase of digital mental health, but also emphasized possibilities (Veldmeijer, Terlouw, Van Os, Van Dijk, Van 't Veer & Boonstra, 2023). For our project, recruiting representatives with double expertise (mental distress experience and AI-development), might overcome (part of) the limitation of common decision-making when it comes to coding, modeling and the mathematical backgrounds of AI. While this might indeed extend the involvement of stakeholders in all phases of AI-research, requesting double-expertise will likely further challenge the representativeness of those involved in co-creation (well-educated, English-speaking, engaged in mental health topics) as compared to the group (very young, vulnerable, limited education) they are representing. To date, it is unclear how this challenge of representing users in all phases of innovative research on digital mental health can be met optimally. How to balance the various needs and possibilities of representing and co-creating is an important topic for further inquiry.

6.4. Recap and future perspectives

We conducted a so-called multilogue with a multidisciplinary team of researchers to point out the human decision making involved in the development of an AI-supported app that aims to support young people and their clinicians in the prediction and attenuation of mental health trajectories. The team covered ethicists, social scientists, psychiatrists, biomedical engineers, AI-developers and young advisors with lived experience. We aim to provide all contributors with equal weight in this approach, despite their different disciplines, roles and status, but must also realize that this is harder to achieve than it might appear. We identified much eagerness and willingness to collaborate and listen to each other's voices, but also different epistemological understandings of mental health, different views on the value of AI for app-development, great engagement in co-creating the appearance and interface of the app but also questions on the possibility to extend meaningful participation to the lowest levels of coding and modeling. Many decisions, however, preceded our scientific work in the consortium and had been taken at a distance and by other actors. In order to even engage in decisions on app-design and the inclusion of AI, all consortium members to some extent complied with set requirements of the EU research policy/grant calls and EU-regulations on data processing. To a certain extent, these boundaries support and structure decision-making, however, they also trigger new ethical questions and lead to ethical dilemmas on good and empowering youth mental health care. To tackle the challenges, we identified during the dialogues and the multilogue, it might be worthwhile to include young people, and a broad variety of scientific disciplines including humanities and social sciences, much earlier in the process: when setting up research calls, during the writing of grant-proposals, in the review procedures and in decisions on awarding grants. In addition, it might be very valuable to also investigate human decision-making, the steps taken and the reasons provided at these preceding and fundamental levels of AI-development. This, however, is well beyond our current purpose.

Despite remaining limits, we conclude that multidisciplinary consortia, the inclusion of a broad variety of disciplines and users in the technical development of AI-supported apps, and also the process of a multilogue itself for explicit reflection upon and exchange of experiences with the ongoing work, are forward-looking and foster constant critical engagements with one's own role in scientific research.

CRediT authorship contribution statement

Dorothee Horstkötter: Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Mariël Kanne:** Writing – review & editing, Methodology, Formal analysis, Data curation, Conceptualization. **Simona Karbouniaris:** Writing – review & editing, Investigation. **Noussair Lazrak:** Writing – review & editing, Investigation. **Maria Bulgheroni:** Writing – review & editing, Investigation. **Ella Sheltawy:** Writing – review & editing, Investigation. **Laura Giani:** Writing – review & editing, Investigation. **Margherita La Gamba:** Writing – review & editing, Investigation. **Esmeralda Ruiz Pujadas:** Writing – review & editing, Investigation. **Marina Camacho:** Writing – review & editing, Investigation. **Finty Royle:** Writing – review & editing, Investigation. **Irene Baggetto:** Writing – review & editing, Investigation. **Sinan Gülöksüz:** Funding acquisition. **Bart Rutten:** Writing – review & editing, Investigation, Funding acquisition. **Jim van Os:** Writing – review & editing, Investigation, Funding acquisition.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding source

Work for this article has been funded by Horizon Europe, grant number 101057182.

References

Aftab, A., & Druss, B. G. (2023). Addressing the mental health crisis in youth—Sick individuals or sick societies? *JAMA Psychiatry, 80*(9), 863–864. https://doi.org/10.1001/jamapsychiatry.2023.1298

Aho, K. (2008). Medicalizing mental health: A phenomenological alternative. *Journal of Medical Humanities, 29*, 243–259.

Amir, T., & Torous, J. (2017). The digital mental health revolution: Opportunities and risks. *Psychiatric Rehabilitation Journal, 40*(3), 263–265.

Arnstein, S. R. (2019). A ladder of citizen participation. *Journal of the American Planning Association, 85*(1), 24–34. https://doi.org/10.1080/01944363.2018.1559388

Arnstein, S. R. (1969). A ladder of citizen participation. *Journal of the American Institute of Planners, 35*, 216–224. https://doi.org/10.1080/01944366908977225

Bak, M. A. R., Ploem, M. C., Tan, H. L., Blom, M. T., & Willems, D. L. (2023). Towards trust-based governance of health data research. *Medicine, Health Care and Philosophy, 26*(2), 185–200. https://doi.org/10.1007/s11019-022-10134-8

Bedor Hiland, E. (2021). *Therapy tech, the digital transformation of healthcare*. Minneapolis: University of Minnesota.

Brey, P., & Dainow, B. (2023). Ethics by design for artificial intelligence. *AI and Ethics*. https://doi.org/10.1007/s43681-023-00330-4

Cary, M. P., Zink, A., Wei, S., Olson, A., Yan, M., Senior, R., & Pencina, M. J. (2023). Mitigating racial and ethnic bias and advancing health equity In clinical algorithms: A scoping review. *Health Affairs, 42*(10), 1359–1368. https://doi.org/10.1377/hlthaff.2023.00553

Correia, A., & Reyes, I. (2020). *AI research and innovation: Europe paving its own way*. Retrieved on 17th June 2024 from: https://research-and-innovation.ec.europa.eu/knowledge-publications-tools-and-data/publications/all-publications/ai-research-and-innovation-europe-paving-its-own-way_en.

d'Aquin, M., Troullinou, P., O'Connor, N.E., Cullen, A., Faller, G., & Holden, L. (2018). *Towards an "Ethics by design" methodology for AI research projects*. Paper presented at the Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, New Orleans, LA, USA. Doi:10.1145/3278721.3278765.

Daley, K., Hungerbuehler, I., Cavanagh, K., Claro, H. G., Swinton, P. A., & Kapps, M. (2020). Preliminary evaluation of the engagement and effectiveness of a mental health chatbot. *Frontiers in Digital Health, 2*. Retrieved on 14th June 2024 from https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2020.576361.

Desmurget, M. (2022). *Screen damage, the dangers of digital media for children*: Wiley.

European Commission. (2023). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Region on a comprehensive approach to mental health*. Retrieved on 23rd September 2024 from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2023%3A298%3AFIN.

European Commission. (2021). Staying healthy 2021, towards a molecular and neurobiological understanding of mental health and mental illness for the benefit of citizens and patients. *HORIZON Research and Innovation Actions*. Retrieved on 9th September 2024 from https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-hlth-2021-stayhlth-01-02.

Foucault M. *The birth of the clinic, an archeology of medical perception*. New York: Pantheon Books, 1973, p. 215.

Fusar-Poli, P., Solmi, M., Brondino, N., Davies, C., Chae, C., Politi, P., & McGuire, P. (2019). Transdiagnostic psychiatry: A systematic review. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA), 18*(2), 192–207. https://doi.org/10.1002/wps.20631

Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard, K. J., & Tørresen, J. (2018). Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing, 51*, 1–26. https://doi.org/10.1016/j.pmcj.2018.09.003

Goisauf, M., & Cano-Abadia, M. (2022). Ethics of AI in radiology: A review of ethical and societal implications. *Frontiers in Big Data, 5*(850383). https://doi.org/10.3389/fdata.2022.850383

Goodman, A., & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child & Adolescent Psychiatry, 48*(4), 400–403. https://doi.org/10.1097/CHI.0b013e3181985068

Grabb, D., Lamparth, M., & Vasan, N. (2024). Risks from language models for automated mental healthcare: Ethics and structure for implementation. *medRxiv*. https://doi.org/10.1101/2024.04.07.24305462

Graham, A. K., Lattie, E. G., & Mohr, D. C. (2019). Experimental therapeutics for digital mental health. *JAMA Psychiatry, 76*(12), 1223–1224. https://doi.org/10.1001/jamapsychiatry.2019.2075

Grote, T., & Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics, 46*(3), 205. https://doi.org/10.1136/medethics-2019-105586

Grote, T., & Berens, P. (2022). How competitors become collaborators—Bridging the gap (s) between machine learning algorithms and clinicians. *Bioethics, 36*(2), 134–142. https://doi.org/10.1111/bioe.12957

Hafner, L., Peifer, T. P., & Hafner, F. S. (2024). Equal accuracy for Andrew and Abubakar—Detecting and mitigating bias in name-ethnicity classification algorithms. *AI & Society, 39*(4), 1605–1629. https://doi.org/10.1007/s00146-022-01619-4

Haidt, J. (2024). *The anxious generation, how the great rewiring of childhood is causing an epidemic of mental illness*: Allan Lane.

Hale, N. (2023). The essence of multilogue, nudges and queries: Enabling un-alinated collaboration space. In J. Baumann Montecinos, T. Grünfelder, & J. Wieland (Eds.), *A relational view on cultural complexities* (pp. 193–213). Cham, CH: Springer.

Ho, A. (2019). Deep Ethical Learning: Taking the interplay of Human and artificial intelligence seriously. *Hastings Center Report, 49*(1), 36–39. https://doi.org/10.1002/hast.977

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L (2018). Artificial intelligence in radiology. *Nature Reviews Cancer, 18*(8), 500–510. https://doi.org/10.1038/s41568-018-0016-5

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kal, D. (2001). *Kwartiermaken, Werken aan ruimte voor mensen met een psychiatrische achtergrond* (p. 213). Amsterdam: Boom [Pathfinding: Working for space for people with a mental health background].

Klengel, T., & Binder, E. B. (2015). Epigenetics of stress-related psychiatric disorders and gene × environment interactions. *Neuron, 86*(6), 1343–1357. https://doi.org/10.1016/j.neuron.2015.05.036

Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., Lenca, P., Billot, R., Ryan, T. C., & Lemey, C. (2021). Machine learning and natural language processing in Mental health: Systematic review. *Journal of Medical Internet Research, 23*(5), Article e15708. https://doi.org/10.2196/15708

Marelli, L., Lievevrouw, E., & Van Hoyweghen, I. (2020). Fit for purpose? The GDPR and the governance of European digital health. *Policy Studies, 41*(5), 447–467. https://doi.org/10.1080/01442872.2020.1724929

Martinez-Martin, N. (2021). Minding the AI: Ethical challenges and practice for AI mental health care tools. In F. Jotterand & M. Ienca (Eds.), *Artificial intelligence in brain and mental health: Philosophical, ethical & policy issues* (pp. 111–125). Cham: Springer International Publishing.

Martinez-Martin, N., Insel, T. R., Dagum, P., Greely, H. T., & Cho, M. K. (2018). Data mining for health: Staking out the ethical territory of digital phenotyping. *npj Digital Medicine, 1*(1), 68. https://doi.org/10.1038/s41746-018-0075-8

Melissa, M., Katrina, H., & Daniel, Z. B. (2023). Evidence, ethics and the promise of artificial intelligence in psychiatry. *Journal of Medical Ethics, 49*(8), 573. https://doi.org/10.1136/jme-2022-108447

Mölders, H., & INCA Projectbureau Amsterdam. (2013). *Multiloog about everyday life*. Paper presented at the *International Conference Psychology and the Conduct of Everyday Life* Roskilde University Copenhagen. Retrieved on 21st June 2024 from https://www.inca-pa.nl/artikelen/multiloog_molders_ipc_roskilde_11072013.pdf.

Moura, I., Teles, A., Viana, D., Marques, J., Coutinho, L., & Silva, F. (2023). Digital phenotying of mental health using multimodal sensing of multiple situations of interest: A Systematic Literature Review. *Journal of Biomedical Informatics, 138*, Article 104278. https://doi.org/10.1016/j.jbi.2022.104278

Musci, R. J., Augustinavicius, J. L., & Volk, H. (2019). Gene-environment interactions in psychiatry: Recent evidence and clinical implications. *Current Psychiatry Reports, 21*(9), 81. https://doi.org/10.1007/s11920-019-1065-5

Nature Mental Health. (2023). Youth mental health crisis management. *Nature Mental Health, 1*(8), 525–526. https://doi.org/10.1038/s44220-023-00112-1

O'Connor, S., & Liu, H. (2024). Gender bias perpetuation and mitigation in AI technologies: Challenges and opportunities. *AI & Society, 39*(4), 2045–2057. https://doi.org/10.1007/s00146-023-01675-4

Renzi, C., Provencal, N., Bassil, K. C., Evers, K., Kihlbom, U., Radford, E. J., & Rutten, B. P. F. (2018). From epigenetic associations to biological and psychosocial explanations in mental health. *Progress in Molecular Biology and Translational Science, 158*, 299–323. https://doi.org/10.1016/bs.pmbts.2018.04.011

Salloch, S., & Eriksen, A. (2024). What are humans doing in the loop? Co-reasoning and practical judgment when using Machine Learning-driven decision aids. *American Journal of Bioethics, 24*(9), 67–78. https://doi.org/10.1080/15265161.2024.2353800

Singh, I., & Rose, N. (2009). Biomarkers in psychiatry. *Nature, 460*, 202–207. https://doi.org/10.1038/460202a

Schubert, W. H. (2008). Problems and possibilities in narrative inquiry, a multilogue. *Thresholds in Education, 34*(1&2), 55–69.

Siala, H., & Wang, Y. (2022). SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. *Social Science & Medicine, 296*, Article 114782. https://doi.org/10.1016/j.socscimed.2022.114782

Smith, K. A., Blease, C., Faurholt-Jepsen, M., Firth, J., Van Daele, T., Moreno, C., & Cipriani, A. (2023). Digital mental health: Challenges and next steps. *BMJ Mental Health, 26*(1), Article e300670. https://doi.org/10.1136/bmjment-2023-300670

Starke, G., De Clercq, E., Borgwardt, S., & Elger, B. S. (2021). Computing schizophrenia: Ethical challenges for machine learning in psychiatry. *Psychological Medicine, 51*(15), 2515–2521. https://doi.org/10.1017/S0033291720001683

United Nations Educational, S., & Organization, C. (2020). *Artificial intelligence and gender equality: Key findings of UNESCO's Global Dialogue*. Retrieved 29th September from https://unesdoc.unesco.org/notice?id=p::usmarcdef_0000374174.

Vargas, C., Whelan, J., Brimblecombe, J., & Allender, S. (2022). Co-creation, co-design, co-production for public health - a perspective on definitions and distinctions. *Public Health Research & Practice, 32*, 2. https://doi.org/10.17061/phrp3222211

Veldmeijer, L., Terlouw, G., Van Os, J., Van Dijk, O., Van 't Veer, J., & Boonstra, N. (2023). The involvement of service users and people with lived experience in mental health care innovation through design: Systematic review. *JMIR Mental Health, 10*, Article e46590. https://doi.org/10.2196/46590

Wolters, P. T. J. (2018). The control by and rights of the data subject under the GDPR. *Journal of Internet Law, 22*(1), 6–18.

World Medical Association. (2024). *Declaration of Helsinki, ethical principles for medical research involving human subjects*. retreived from https://www.wma.net/policies-post/wma-declaration-of-helsinki/.