Building Responsible AI for Mental Health: Insights from the First RAI4MH Workshop

White paper

NOVEMBER 6TH, 2024

University of Southampton
Institute for Experiential AI @ Northeastern University

Authored by:

Dr Rafael Mestre
Dr Annika M. Schoene
Dr Stuart E. Middleton
Prof Agata Lapedriza

UNIVERSITY OF Southampton

EAI The Institute for Experiential AI
Northeastern University

RAi

With the support of:

THE CENTRE FOR DEMOCRATIC FUTURES

Centre For MachineIntelligence

Cite as:

Mestre, R., Schoene, A. M., Middleton, S. E., & Lapedriza, A. (2024). *Building Responsible AI for Mental Health: Insights from the First RAI4MH Workshop* [White paper]. University of Southampton and Institute for Experiential AI @ Northeastern University. https://doi.org/10.5281/zenodo.14044362

The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the University of Southampton or the Institute for Experiential AI at Northeastern University.

Responsible AI for Mental Health (RAI4MH) partnership
www.rai4mh.com
rai4mh@soton.ac.uk

About the authors

Rafael Mestre

Dr Rafael Mestre is a Lecturer at the University of Southampton and a Turing Fellow at the Alan Turing Institute. His research is interdisciplinary and focuses on the responsible development, socio-technical evaluation and co-governance of emerging technologies, at the interface of disciplines like computer science, science and technology studies, health and political science. He has worked in and studied emerging technologies like AI (particularly multimodal machine learning), biohybrid robotics, distributed acoustic sensing, tissue engineering and 3D bioprinting, focusing on their applications in political science, robotics, smart cities and health.

Annika M. Schoene

Dr Annika M. Schoene is a Research Scientist at the Institute of Experiential AI (EAI) in the AI + Health Team and Responsible AI Practice. Her work investigates how social, economic, political, and physical factors are driving forces behind health inequalities using AI. More specifically, she focuses on Social Determinants of Health using Natural Language Processing and their impact on onset and progression of mental health conditions and substance use disorders. She also gathers evidence for non-technical stakeholders and policy makers to make informed decisions that reduce algorithmic harm in healthcare settings.

Stuart E. Middleton

Dr Stuart E. Middleton is an Associate Professor at the University of Southampton. He has more than 60 peer reviewed publications, many inter-disciplinary in nature, focussing on the Natural Language Processing (NLP) areas of information extraction and human-in-the-loop NLP. His research interests are focussed on socio-technical NLP approaches, including large language models, few/zero-shot learning, rationale-based learning, adversarial training and argument mining. He has worked in domains including law enforcement, defence and security, mental health, environmental science, legal and misinformation.

Agata Lapedriza

Prof Agata Lapedriza is a Principal Research Scientist at the Institute for Experiential AI, an Affiliated Professor at Bouvé College of Computing at Northeastern University, and a Professor at Universitat Oberta de Catalunya, with a background in Computer Vision and Machine Learning. Her current research focuses on Human-Centric AI, particularly on computational models to analyse, recognise and interpret human-related aspects from visual data, language, and data captured with wearable sensors. Lapedriza has been a research affiliate at MIT since 2012.

About RAI4MH

In recent years, the growing number of individuals facing mental health challenges is placing an unprecedented strain on our healthcare services to facilitate adequate care. At the same time, Artificial Intelligence (AI) has increasingly been used as a supplement for triaging, diagnosis and treatment recommendations both in public and private healthcare settings.

Our project is an international collaboration between the **University of Southampton** in the UK and the **Institute for Experiential AI at Northeastern University** (USA), aiming to lead the conversation on how AI can be used responsibly in mental health care. We bring together experts from mental health, computer science, ethics, and policy, along with voices from industry, the third sector, and public sector, with the goal to create guidelines that ensure user protection, privacy, and fairness. These guidelines will help in applying AI responsibly within health services, potentially easing the burden on healthcare providers and enhancing patient care. We are keen to engage with stakeholders from all sectors – from public agencies to private industry – to gather insights and shape these guidelines effectively.

www.rai4mh.com

rai4mh@soton.ac.uk

Executive Summary

Introduction

The **Responsible AI for Mental Health (RAI4MH) workshop**, held in London, convened over 65 experts in person (and more than 130 online) from academia, industry, public sector, and healthcare to explore the ethical and practical challenges of integrating AI into mental health care. As mental health needs escalate worldwide, Artificial Intelligence (AI) is increasingly seen as a tool to enhance access, diagnostics, and support. However, given the high stakes of mental health care, the responsible application of AI requires strong ethical frameworks, guidelines, and transparent governance.

Workshop overview

The two-day hybrid event was structured to balance broad discussions with focused, actionable outcomes. On Day 1, participants engaged in thematic discussions and panel sessions, with speakers outlining topics from AI's potential to expand mental health services to the ethical considerations necessary to avoid misuse, misinterpretation, and privacy breaches. Day 2 transitioned to small group work, where 24 selected experts used structured exercises, including grid analysis and scenario modelling, to identify core challenges and develop practical solutions. This intensive session led to a set of preliminary policy recommendations, focusing on the essential conditions and practices needed for responsible AI deployment in mental health.

Insights and strategic priorities

Participants highlighted AI's potential to improve accessibility and early intervention in mental health, offering timely support, enhanced diagnostics, and broader access across groups. Aspirations included AI-enabled diagnostic accuracy, scalable support, and cost-effectiveness while preserving patient autonomy and human connection. Key challenges emerged around infrastructure limitations, data privacy, and transparency, with concerns about biases, dependency on AI, and accountability gaps. Governance structures, robust ethical standards, and data protections were seen as essential to build public trust and confidence, with regulation viewed as a facilitator of innovation rather than a constraint.

Preliminary policy recommendations

The workshop's preliminary policy recommendations emphasise building a robust support system for AI in mental healthcare through high-impact actions like funding healthcare infrastructure, upskilling the workforce, and ensuring data security via digital onshoring and localised healthcare data centres. Other key measures include monitoring AI's long-term effects through regular check-ins, supporting ethical integration with clear pathways for AI use, and offering specific guidelines for Software as Medical Device (SAMD). Medium-effort recommendations aim to establish ethical standards for AI, especially for at-risk individuals, while low-effort actions like fostering public understanding, interdisciplinary collaboration, and sharing an evidence base promote trust and enhance AI literacy.

1. Introduction

The world is grappling with an escalating mental health crisis, especially among young people (Daly, 2022). Factors such as societal pressures, economic instability, and the unique stressors introduced by the digital age contribute to this worsening crisis, straining existing mental health infrastructures and leaving millions without adequate care and support. In the UK, around 1.5M people are currently accessing secondary mental health services (UK Government, 2024). Simultaneously, the shortage of mental health professionals and resources highlights an urgent need for innovative solutions to bridge this gap.

With mental health needs far outpacing available resources, emerging technologies like Artificial Intelligence (AI)—and specifically Natural Language Processing (NLP)—offer potential tools to address mental health challenges. Across the health sector, AI applications are proposed in triaging, diagnosis, and prognosis, allowing medical professionals to allocate resources more efficiently and respond to cases with greater accuracy and speed. For example, AI-driven tools could analyse vast amounts of clinical and patient data to detect patterns that may indicate early signs of mental health issues, making it possible to identify and address issues like online abuse, bullying, self-harm tendencies, and suicide ideation in real-time (Zirikly, 2022; Azim, 2022; Schoene, 2022, 2023, 2024). Yet, applying AI within mental health care must be approached with caution due to the high stakes of misuse, misinterpretation, and the potential for exacerbating existing health disparities.

At the heart of responsible AI (RAI) in mental health is the need for ethical consideration and safety, fairness, and transparency in every AI deployment (Jobin, Ienca, & Vayena, 2019). The integration of AI within mental health care is filled with challenges and must be approached with caution. AI algorithms are inherently complex and can be prone to biases, misinterpretations, or misuse, potentially leading to adverse outcomes if not implemented carefully. Robust and responsible evaluation and the involvement of interdisciplinary teams—including clinicians, researchers, technologists, and policymakers—are critical to shaping AI tools that reflect diverse perspectives and address real-world mental health needs.

This white paper explores these challenges and synthesises the insights from a two-day workshop with diverse stakeholders, aimed at developing policy recommendations for responsible AI use in mental health.

2. Workshop overview

The **Responsible AI for Mental Health (RAI4MH)** workshop, held in London, was a two-day hybrid event that engaged a broad range of experts from academia, industry, government, and the non-profit sector to explore AI's role in mental health. On the first day, we welcomed more than 65 people in person and over 130 attendees online, including 15 international speakers and panellists[1].

[1] Recordings of the day can be found on YouTube: https://youtu.be/zo-8Uhnu90Q.

Through keynotes, panels, and interactive hybrid Q&A sessions, the event aimed to surface both the promise and ethical challenges of AI in mental health.

On the second day, a smaller group of 24 experts met in a focused workshop designed to translate Day 1's discussions into practical recommendations. Using a collaborative, structured process, participants engaged in exercises such as grid analysis and scenario modelling to identify key areas where policy could guide responsible AI integration in mental health. This smaller group represented diverse expertise, with participants from psychology, law, computer science, and public health policy, encouraging a holistic view of the challenges and opportunities AI presents in mental health.

Day 1: Expert talks and foundational discussions

The workshop's first day provided a foundation through thematic sessions and panel discussions. After opening remarks by co-organiser **Dr. Rafael Mestre**, our morning session on *Digital Mental Health* was kicked off by **Dr. Aynsley Bernard** from Kooth Plc., a leading digital mental health provider that offers online counselling and support services for young people and adults. She discussed Kooth's responsible innovation practices in digital mental health. This was followed by a panel featuring **Stuart Pearson** (Citizens Advice), **Dr. Becky Inkster** (University of Cambridge and self-employed), **Dr. Daniel Leightley** (King's College London), and **Nick Pollard** (FamilyMentalWealth), moderated by **Prof. Agata Lapedriza** (Northeastern University and Open University of Catalonia). They discussed the future of digital mental health, emphasising the need for collaboration across fields for responsible development. To conclude the session, **Prof. Elvira Pérez-Vallejos** from the University of Nottingham spoke about the application of digital tools for death and aging, highlighting important insights in this often-overlooked aspect of mental health care.

The afternoon session on *Responsible Use of AI in Mental Health* started with an overview of responsible AI by **Prof. Ricardo Baeza-Yates**, Director of Research at the Institute for Experiential AI at Northeastern University. This was followed by **Dr. Aminat Adebiyi** from IBM Research, who joined us online to discuss the importance of social value alignment in AI and IBM's efforts in responsible AI. **Prof. Maria Liakata** from Queen Mary University of London then spoke about the socio-technical limitations of large language models (LLMs) for social computing. The day ended with a panel featuring **Dr. Laura Haaber Ihle** (Northeastern University), **Dr. Joseph Connor** (CarefulAI), **Mariana Silva** (University of Cambridge), and **Dr. Brieuc Lehmann** (University College London), moderated by **Dr. Stuart E. Middleton** from the University of Southampton. They discussed the challenges in applying AI responsibly in the mental health sector, including ethical considerations and data-driven tools. Co-organiser **Dr. Annika Marie Schoene** closed the event with some thoughtful remarks.

These sessions were combined with opportunities for participants to engage directly through polls and Q&As, allowing them to voice their perspectives on AI's ethical implications and highlight specific areas of concern or optimism. This input provided a foundation for more targeted discussions on the second day.

Day 2: Collaborative policy development

On the second day, a smaller, focused group of 24 participants worked through structured exercises to draft policy recommendations around AI use in mental health. The workshop was designed following a Double Diamond design process (Kochanowska & Gagliardi, 2022), commonly used to tackle complex problems, with the aim of converging into preliminary policy recommendations that could tackle the variety of issues highlighted on the day before. The day was divided into four phases.

Discover phase: Grid Analysis

In the ‘discover’ phase of the workshop, participants engaged in grid analysis to explore the sector’s core Opportunities, Aspirations, Challenges, and Risks (OACR), divided into three groups. For that, they were given a series of seed prompts to start their discussions in the form of *post-it* notes. These were a series of key topics that reflected the overall conversation taking place on the day before, including also clustered key topics that attendees highlighted in the online poll the day before. These ideas could be positive or negative, ranging from issues like “accessibility,” “infrastructure” and “data privacy” to “better triage” and “earlier interventions.”

Define phase: Prioritisation

In the ‘define’ phase, participants were asked to vote and prioritise the most important topics, individually, from the outputs of all groups. This allowed them to identify the key issues, whether they were opportunities or risks, that were salient and considered more important across all groups.

Develop phase: Scenario modelling

In the ‘develop phase’, participants were asked to brainstorm and start coming up with potential solutions to tackle the issues that received the most votes. Working in groups, they created scenarios to illustrate how AI might function in real-world mental health applications, allowing them to explore potential ethical challenges, user outcomes, and broader implications for public trust. Using both positive or negative priorities from the previous activities (e.g., opportunities or risks), participants thought through complex issues—such as the balance between automated and human-centred care, the impact of data governance, and accessibility limitations—to ensure that their policy recommendations accounted for a range of potential user experiences and ethical considerations.

Deliver phase: Drafting recommendations

Finally, in the ‘deliver’ phase, participants were asked to converge in policy recommendations that would help solve those issues and maximise positive outcomes. They were asked to place them in a grid of ‘impact vs effort’. Effort was broadly undefined so participants could self-define what effort meant in terms of ‘monetary cost’, ‘time’, ‘feasibility’, ‘human cost’, etc.

3. Insights and strategic priorities

In the ‘discover’ phase with grid analysis, participants highlighted was AI’s potential to improve **accessibility and early intervention** in mental health. Participants emphasised that AI could deliver timely, scalable support for mental health care, offering quicker responses and more consistent service quality. Additionally, AI was seen as a powerful tool for **enhanced diagnostics and triage**; it could enable improved diagnostic accuracy by tracking symptoms or analysing patient health histories over time. Digital interventions also emerged as a key opportunity, where AI could support various forms of therapeutic interactions, expanding access to support across different population groups and regions.

Participants also expressed aspirations for AI-enhanced mental health system marked by **compelling evidence of efficacy** in both health outcomes and cost-effectiveness. They also highlighted aspirations for **better diagnostic** and **triage capabilities**, which could support more timely and effective mental health interventions, as well as hopes for AI to drive **greater accessibility** to diverse populations while adhering to **ethical guidelines.** This includes preserving **patient autonomy** and upholding values like **authentic human connection** and **dignity** in treatment. Transparency in AI model function and metadata was emphasised to ensure that AI aligns with both user and ethical expectations.

They identified several key challenges that must be addressed for AI to effectively support mental health care. A primary concern was the **infrastructure gap**, especially within systems like the NHS, where there may be limitations in technology and digital literacy among both providers and users. Additionally, participants noted the need for **better data quality** to ensure that AI models are accurate and reliable, as well as the necessity of balancing **business priorities** with ethical considerations in health-focused AI initiatives. **Data privacy and governance** also emerged as critical concerns, as well as **intersectional accountability** to prevent biases and ensure AI tools serve diverse communities equitably. Other challenges included **environmental sustainability**, and building a strong **evidence base** for AI efficacy in terms of both cost and mental health outcomes.

Participants noted several risks associated with AI applications in mental health, with **trust and transparency** standing out as critical. There was concern that insufficient transparency in AI operations could undermine public trust and reduce the perceived reliability of AI-driven mental health tools. Additionally, participants highlighted the risk of **emotional dependency** on AI, where users might over-rely on AI tools, potentially compromising genuine human connection and individualised care. The potential **lack of personalisation** was also seen as a risk, as AI-driven solutions may struggle to address the unique needs of each patient, affecting treatment quality. Issues of **accountability** and **critical thinking** were noted, particularly around the potential reluctance to report AI system failures or

Table 1. Most voted topics (with at least 3 votes) in the prioritisation phase.

Issue	Votes	Category
RAI governance	7	**Necessary Prerequisites**
Evidence basis (efficacy, cost/benefit)	6	**Challenges**
AI delivery (scalability, early/timely intervention, quicker response, multimodal support, consistency)	5	**Opportunities**
Better triage	5	**Aspirations**
Intersectional accountability	5	**Challenges**
Preserving authentic human connection	5	**Aspirations**
Better mental health for more people	5	**Aspirations**
Accountability	4	**Risks**
Infrastructure (e.g., NHS)	4	**Challenges**
Responsible data governance	4	**Challenges**
Transparency / Lack of understanding	3	**Necessary prerequisites**
Lack of critical thinking around use of AI	3	**Risks**
Public acceptability / trust	3	**Outcomes**
(Un)willingness to report failure	3	**Risks**
Earlier intervention	3	**Opportunities**

inaccuracies. Finally, participants warned about **data commodification** and **loss of confidentiality**, where sensitive mental health data could be misused or exploited, leading to privacy breaches and ethical concerns regarding the commercialisation of mental health information.

One group in the workshop expanded on the standard categories by introducing two additional ones: **Necessary Prerequisites** and **Outcomes.** These categories were meant to reflect the sense that responsible AI in mental health requires a strong ethical and practical foundation, alongside clear, measurable goals for public impact. In the **Necessary Prerequisites** category, participants identified core requirements, such as robust governance structures, accountability mechanisms, and data privacy protections. They emphasised that without these foundational elements, other goals would be difficult to achieve. Transparency, reducing model biases, managing epistemic risks (such as misinterpretation of data), and securing dedicated research funding were seen as essential first steps toward responsible AI deployment. For **Outcomes**, the group focused on long-term goals for successful AI integration, particularly **public acceptability,** and **trust.** Rather than seeing trust as an end goal, participants viewed it as an organic outcome that follows from transparent, ethical, and inclusive AI practices. They highlighted that achieving these outcomes requires building confidence in AI through transparency, ethical standards, and inclusive development practices.

The prioritisation activity (**Table 1**) highlighted a balanced view of both the opportunities and challenges of using AI in mental health, showing that participants were not solely focused on risks but also saw significant potential for positive impact and aspirations. The top-ranked issue, **RAI governance**, with 7 votes, highlights the priority given to establishing clear policies, standards, and oversight mechanisms to ensure responsible AI practices. This focus on governance was not seen as a barrier but rather as a necessary foundation to unlock the potential of AI, reflected in the high ranking of opportunities and aspirations like improving **AI delivery** (scalability and early interventions) and supporting **earlier interventions**.

Aspirations for **better triage** and **reaching more people** suggest optimism about AI's ability to enhance care, but with the important caveat that it should preserve **authentic human connection**—a theme that was consistently highlighted on both days of the event. This balance of optimism and caution shows that regulation and governance are not seen as roadblocks but as essential enablers of innovation. Risks, such as **accountability gaps** and **lack of critical thinking**, were acknowledged, but they did not dominate the discussion. Instead, they were considered alongside opportunities and aspirations, suggesting that participants view regulation not as a restriction but as a tool to guide AI's positive potential in mental health.

4. Preliminary policy recommendations

The workshop's preliminary policy recommendations (**Table 2**) capture the collective expertise and dedication of its participants, who brought together perspectives from academia, industry, public policy, and healthcare. Out of those with an estimated high-impact, high-effort recommendations, such as **funding healthcare infrastructure** and **upskilling the workforce**, aimed to build a strong support system for AI deployment, both in terms of human and infrastructure capital. These came accompanied by **monitoring negative effects** and **deliberate care management**, which proposed periodic check-ins to ensure people redirected to mental health tools or services were using them and working in the appropriate way, not only at the beginning of the care pathway. Others, such as **digital onshoring** and **creating foundations for healthcare data centres** highlight the need to bring data and digital processes to common local regions, which would bring benefits to the economy and allow better control of sensitive data. Despite their high effort, those were seen as critical recommendations to ensure a sustainable, responsible and robust AI deployment in healthcare and mental healthcare in particular.

Medium-effort actions, including **pre-planned and defined pathways**, focus on ensuring structured, ethical integration of AI, especially when dealing with individuals at risk of harm to others or themselves. Low-effort recommendations, like **clear guidance for Software as Medical Device**

Table 2. Preliminary policy recommendations by expert participants ranked in an impact-effort scale.

Impact / Effort	High Effort	Medium Effort	Low Effort
High Impact	- Funding infrastructure (to ensure the end of IT monopolies in healthcare) - Upskilling workforce - Digital onshoring (benefit to economy) - Create foundations for data centres (healthcare) - Monitoring negative effects - Deliberate care management (check-ins in patients using tools)	- Pre-planned and defined pathways as standard for dealing with AI-assisted mental health - Clarity on pathways for standardisation, regulation, and evidence generation - Support professionals in AI procurement (public sector)	- Clear guidelines for Software As Medical Device (SAMD) - Support development of evidence basis (sharing fails, etc.)
Medium Impact	- Continuity of data and portability in the care pathway	- Value alignment (relevant to culture and context) - Create public and free governance structure across different regulators - Have regulators share protocols, trainings, examples, reusable components, etc. - Mirror EU law/legal aspects - Recognise hierarchies of risks and how these intersect (equalities impact assessment) - Give monitoring & regulatory power enforcement	- Public engagement - Build skills in education - Interdisciplinary collaboration - Translating standards for non-technical audiences - Recognise AI in relation to physical services / support
Low Impact	- Build understanding of AI in public sector		

(SAMD) and support the **development of evidence basis**, such as sharing fails within the community, provide quick wins for building trust and establishing a reliable foundation, which could be integrated with other medium-impact activities like **public engagement**, **building skills**, or **enabling interdisciplinary collaborations.**

A series of activities related to education, training and mindset-shifting like **building skills in education**, **improving public understanding**, **translating standards for non-technical audiences** and **enabling interdisciplinary collaborations** were seen as low effort but with the potential for medium impact. These came accompanied by the need **to recognise AI in relation to physical services and support**, as physical spaces tend to be confused as a different space away from humans. With the potential of medium impact, but of medium effort, came a series of recommendations around regulations and governance structures, with a clear acknowledgement of the importance of **value alignment** and **impact on different communities**.

Finally, some recommendations were considered to require a medium-to-high effort, but a lower impact (although not insignificant) when compared to the rest of them. These were **ensuring the continuity of data and its portability in the care pathway** of the patient, especially when dealing with different doctors, providers and processes, and to **build the public understanding of AI in the public sector**.

5. Conclusions

The discussions and work carried out at the RAI4MH workshop underscored both the potential and the complexities involved in responsibly integrating AI into mental health care. Participants highlighted that AI, if properly governed and ethically implemented, could enhance access, improve early intervention, and provide scalable support for mental health, offering new avenues to address the current crisis. However, this potential is contingent upon establishing strong ethical frameworks and foundational prerequisites such as robust governance structures, data privacy protections, and transparency measures.

A key takeaway from the workshop was the **balanced perspective** that participants maintained: **regulation and ethical guidelines were not seen as barriers to innovation but as enablers** that can build public trust and ensure AI's alignment with patient-centred values. Participants identified recommendations for both immediate needs and long-term systemic changes.

To maximise impact, high-priority recommendations focused on **building a strong support system** for AI integration in healthcare through initiatives like **funding infrastructure** (like data centres), **digital onshoring** and **upskilling the workforce**, which will enhance healthcare resilience by fortifying both human and technological resources, and ensure secure, localised data management. **Monitoring the long-term effects of AI-driven interventions** through periodic patient check-ins and very clear guidelines and pathways also emerged as essential recommendations, as mental health tools and services should maintain effectiveness and patient safety over time. **Structured pathways for ethical AI integration** and clarity in guidelines, such as for Software as Medical Device (SAMD), were seen as pivotal for supporting safe and consistent AI implementation in mental healthcare.

Additionally, efforts focused on education, public engagement, and skill-building, though lower in effort, were recognised as having significant potential for cultural shifts. Collectively, all of these recommendations have the potential to lay a solid foundation for a secure, responsible, and equitable approach to AI in healthcare, meeting ethical standards while supporting patient well-being and systemic resilience.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/Y009800/1], through funding from Responsible Ai UK (IP0033). The authors would like to thank Kooth Plc., partners in this initiative, as well as all the keynote speakers and panellists who lent their time to make this event possible, and all our attendees, both in person and online. Recordings of the first day can be found on YouTube: https://youtu.be/zo-8Uhnu90Q. We wholeheartedly thank those who attended the workshop on the second day and gave their input to shape the policy recommendations: Aislinn Gomez Bergin, Aynsley Bernard, Becky Inkster, Brian Tschuma, Daniel Leightley, Elvira Perez Vallejos, Emma Palmer-Cooper, Gyanendro Loitongbam, Hannah Gardiner, Lana Bojanic, Laura Haaber Ihle, Mariana Silva, Melika Vafafar, Melyssa Ortiz, Ricardo Baeza-Yates, Richard Gomer, Sarah Kiden, Sarah Lewthwaite, Sarah Jenner and Sergey Astakhov.

References

Azim, T. Singh, L. Middleton, S.E. Detecting Moments of Change and Suicidal Risks in Longitudinal User Texts Using Multi-task Learning, In Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology, NAACL-2022, pages 213–218, Seattle, USA. Association for Computational Linguistics, 2022.

Daly, M., Sutin, A., & Robinson, E. (2022). Longitudinal changes in mental health and the COVID-19 pandemic: Evidence from the UK Household Longitudinal Study. Psychological Medicine, 52(13), 2549-2558. doi:10.1017/S0033291720004432.

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

Kochanowska, M., & Gagliardi, W. R. (2022). The double diamond model: In pursuit of simplicity and flexibility. *Perspectives on Design II: Research, Education and Practice*, 19-32.

UK Government. (2024, October 24). *Mental health services monthly statistics performance, August 2024*. GOV.UK. https://www.gov.uk/government/statistics/mental-health-services-monthly-statistics-performance-august-2024.

Schoene, A.M., Ortega, J., Amir, S. and Church, K., 2023, June. An Example of (Too Much) Hyper-Parameter Tuning In Suicide Ideation Detection. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 17, pp. 1158-1162).

Schoene, A.M., Bojanić, L., Nghiem, M.Q., Hunt, I.M. and Ananiadou, S., 2022. Classifying suicide-related content and emotions on Twitter using Graph Convolutional Neural Networks. IEEE Transactions on Affective Computing.

Zirikly, A. Atzil-Slonim, D. Liakata, M. Bedrick, S. Desmet, B. Ireland, M. Lee, A. MacAvaney, S. Purver, M. Resnik, R. Yates, A.. 2022. Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. Association for Computational Linguistics, Seattle, USA, edition.