NCPS | NATIONAL COUNSELLING & PSYCHOTHERAPY SOCIETY

National Commission into the Regulation of AI in Healthcare: Call for Evidence

Section One – Respondent Information

Question 1: Are you responding as an individual or on behalf of an organisation?

- Organisation

Question 3.1: What is the name of the organisation you are representing?

- National Counselling & Psychotherapy Society

Question 3.2: What type of organisation is this?

- Other – Professional body for counsellors & psychotherapists

Question 3.2.1: Which of the following best describes your organisation?

- My organisation does not develop healthcare AI products

Question 4: We may want to follow up with you – if you are happy to be contacted, please provide us with a contact name, organisation (if relevant) and email address.

- Meg Moss, Head of Public Affairs & Advocacy, NCPS: meg@ncps.com

Section Two – Call for Evidence Questions

Question 1: Which of the following best describes your view about the need to change the UK's framework for regulating AI in healthcare?

- Significant reform: The current framework requires substantial changes

Question 4: How might the UK's framework for regulation of AI in healthcare be improved to ensure that NHS has fast access to safe and effective AI health technology?

The risk levels in the framework for regulating artificial intelligence should be stratified, with mental health and psychological support recognised as a higher-risk area of application. Mental health and wellbeing are shaped through complex interpersonal, social, and relational processes, and support in this area often involves working directly with people who are vulnerable, or perhaps experiencing issues related to identity, or emotional regulation. AI systems used in this context therefore will clearly raise risks that differ in both nature and timescale from those associated with, for example, diagnostics or administration. An AI tool used to summarise clinical notes presents a very different risk profile from a system that responds directly to someone expressing distress, loneliness, or trauma, and a different risk profile again from an AI tool used to understand the implications of someone's blood pressure results.

We would like to see greater clarity around the classification of AI systems that engage with people who are experiencing emotional distress, or psychological meaning-making, or any kind of therapeutic-style interaction. Many of these tools operate in practice within grey areas, presenting themselves as wellbeing support while being experienced by users as therapeutic. So, this means that a system marketed as providing general mental wellbeing advice may, in reality, be used by people as their primary source of emotional support, particularly where access to human services is limited, either due to lengthy waiting times or the type of support needed being unavailable in that area. Regulation should take account of how systems are actually used and relied on in reality, rather than focusing only on what their stated intended purpose is.

We would also like to see evidence requirements being strengthened for mental health applications. Assessment measures based primarily on short-term engagement or symptom change risk missing the bigger picture; for example, high levels of repeated use may indicate that the person is benefitting from the service, but it may also suggest dependency or a lack of onward support. Evidence should therefore include longer-term and qualitative outcomes, such as whether people using AI support show sustained improvement, increased resilience, or appropriate engagement with human services. It should also explore whether AI use is associated with disengagement from the service; simply assuming that because a person is no longer using the service, they no longer require it, can no longer be an acceptable outcome. It will be imperative to ascertain whether or not AI tools have become a barrier to accessing a service, as we know that attrition and dropout for app-based health services is high (https://pmc.ncbi.nlm.nih.gov/articles/PMC7556375/ / https://pmc.ncbi.nlm.nih.gov/articles/PMC11694054/).

From the perspective of a professional body representing counsellors & psychotherapists, our profession is built on the knowledge that effective mental health support depends on the gradual building of trust, safety, and appropriate emotional containment within a relationship, which means recognising that meaningful change is often the result of consistent, boundaried human relationship, and can't be replaced by rapid or transactional interactions. Ethical practice in mental health is closely tied to the quality of the relationship through which therapy is provided, and this should be reflected in how AI tools are assessed and regulated to determine how they support, and don't detract from, human-led provision.

We also have concerns about disinhibition in digital interactions, where people may disclose highly personal or distressing information more quickly when engaging with AI systems, particularly where there is no perceived judgement or consequence. For example, someone may share traumatic experiences or express thoughts of self-harm earlier than they would in a human relationship, without the presence of a professional who can appropriately contain or respond to that disclosure. Without appropriate relational holding, this rapid unburdening may leave people emotionally exposed or unsettled, even if the interaction feels supportive in the moment. The safeguarding protocols in those situations should be clear, and how levels of risk are assessed in order to enact those safeguarding protocols should be clearly considered and outlined.

Clearer guidance is needed for NHS commissioners and healthcare providers on how AI tools should be used within mental health pathways. Where they are introduced, these tools should complement human support rather than replace it. For example, AI might support signposting, appointment navigation, or administrative tasks, but should not function as a substitute for therapeutic relationships. While timely access to help is important, speed should not come at the expense of careful oversight in areas where harm may be subtle and difficult to identify, and potentially worsening as time goes on as the cumulation of smaller therapeutic harms compound. One example already in use within healthcare settings is the use of AI to triage patients into the service; however, what is currently unclear is how this tool can account for the natural nuance, and often dissonant ways in which people first present to mental health services. As humans, we're well aware when a person's body language and presentation are communicating something different to their words; how do we seek to understand this in the context of a service that has no human involvement, relying solely on the words typed or boxes ticked?

So, to sum up, we would like to see a regulatory framework for artificial intelligence in healthcare that clearly distinguishes between levels of risk for different applications, either where the potential for harm is greater, or less visible / easily quantified. This should include clearer classification and understanding of the risks, stronger and more appropriate evidence requirements, and explicit safeguards to ensure AI tools are used to support, rather than replace, therapeutic human relationships. If this can be achieved, regulators and commissioners will be in a much stronger position to determine which AI interventions can be deployed safely and at pace, and which require more time, oversight, and consideration before being introduced into mental health support pathways.

Question 5: How should the regulatory framework manage post-market surveillance for AI health technologies?
Post-market surveillance in terms of usage in mental health care in particular should be continuous, proportionate to risk, and responsive to how systems are used in practice.

Harms may not always be clear-cut or tangible discrete incidents (although of course they may be, as we have seen with a number of chatbot-related deaths), and we may see them develop gradually through relational factors like withdrawal from human relationships, reduced friction-tolerance, reduced distress-tolerance; or perhaps changes in how the service is used that may show an overreliance on the service, or even disengagement from mental health support in general that could potentially indicate a lack of connection with, or trust in, the mental health support service provided. Where people do stop using an AI-supported service, efforts should be made to understand why. For example, have they transitioned to another NHS service, sought support outside the NHS, or disengaged from support altogether?

For AI systems used in mental health contexts in particular, surveillance should include active monitoring of use beyond the system's stated intended purpose. People tend to use chatbots in ways in which they were not originally designed; for example, using them as their primary source for emotional support, rather than as a supplementary or signposting tool.

Measures will also need to cover psychological and relational harms, which should include concerns such as increased dependency, emotional distress linked to use of AI tools, or evidence of delayed or reduced engagement with human support.

Qualitative user feedback will be vital in understanding how risks are borne out, and not just for those users that continue to engage with support, but those who have dropped out of the system as well.
It's also important to consider the wider system context; for example, monitoring should include how AI use correlates with staffing levels of human practitioners within individual NHS trusts, and whether AI is being introduced primarily as a cost-saving measure rather than as a clinically appropriate adjunct. We would also like to see it examine the impact of AI deployment on workforce capacity, professional roles, and the availability of relational support, particularly in services already under pressure.

The framework, or supplementary guidance, should include a requirement for clear escalation mechanisms where concerns are identified. These should include the ability to modify, suspend, or withdraw AI systems where evidence suggests they are contributing to harm or being used inappropriately.
All of this should be underpinned by robust, long-term research, supported by the regulator, which includes studies examining the quality of relationships over time, levels of wellbeing and resilience, and broader social outcomes, while accounting for other influencing factors. The idea is to understand not just whether AI systems are safe in the short term, but whether they contribute positively to mental wellbeing over the longer term, and to understand more fully what the relational impacts are of integrating non-human relationships into human-centred work.

Question 7: How could manufacturers of AI health technologies, healthcare provider organisations, healthcare professionals, and other parties best share responsibility for ensuring AI is used safely and responsibly?

Manufacturers of AI health technologies should be responsible for the ethical design, development, and ongoing evaluation of their systems. This includes being transparent about how tools function, what assumptions, models, or therapeutic approaches are embedded within them, and what their limitations are. AI systems may reflect particular psychological models, cultural assumptions, or biases that aren't visible to end users or clinicians, so manufacturers should engage proactively with professional bodies, service-user groups, and experts during the design stage to make sure their tools are appropriate for use.

Healthcare providers and commissioners have ultimate responsibility for deciding how and where AI tools are deployed, which includes ensuring that technologies are commissioned only for uses that are clinically appropriate and ethically justified – considerations around the scale of the deployment, the scope, and where and how human interaction or oversight is replaced all sit at this level.

Clinicians / practitioners should be responsible for using AI tools appropriately within the scope defined by their role and training. Clinicians should be given adequate, structured training by their employer or the commissioner, which would include clear guidance on when AI tools should and should not be used. This is particularly important where AI might be used for tasks such as note-taking, summarising sessions, generating suggested treatment plans, or processing highly sensitive client data. Clinicians must be able to understand the limits of these tools, the risks involved, and the implications for confidentiality, consent, and the therapeutic relationship. It is incumbent on the commissioner to provide good quality training that explains the breadth of risk involved, including relational risks, and how they might be mitigated. It's also incumbent on the commissioner to ensure the service they are providing is appropriate, and any issues that arise due to the inappropriateness of the service in general are not the responsibility of clinicians.

Transparency is essential at all levels; clinicians and service users should be able to understand when AI is being used, for what purpose, and with what safeguards in place, which includes transparency about data use, the limitations of the model, the modality on which the model is based, the extent to which outputs are automated or human-led, and any guidance around recommended or safe usage. It should be clear where the data is being stored, and whether or not it is being used to further train the model. It should be clear about the ways in which it will be used to make decisions about patients, and there should be clear accountability for how clinical decisions are made and reviewed by AI systems; there should always be a human-in-the-loop.

It's important that all parties involved remain engaged in the process of continually understanding, monitoring, and developing the tools or services, and that candour and openness remain central tenets of any safety framework.

Question 8: in the event of an adverse patient outcome where and adverse patient outcomes involved an AI tool, where do you think liability should lie?

Those who have control, decision-making authority, and foreseeability of risk should take liability in most cases, which means that responsibility likely rests with manufacturers and the organisations that choose to deploy and commission AI systems, rather than with individual clinicians or users. Manufacturers should be liable where harm arises from system design, training data, embedded assumptions, or known limitations that were foreseeable but not adequately mitigated or communicated. This includes situations where AI tools encode particular models of mental health support, introduce bias, or function in ways that are not transparent to clinicians or service users. Where a system behaves in ways that are consistent with its design but nevertheless cause harm in real-world use, liability should not be transferred to those who did not design or control the system.

Deploying organisations and commissioners should bear responsibility for decisions about where, how, and at what scale AI tools are used. This includes situations where AI is introduced into mental health pathways inappropriately, used beyond its intended purpose, or substituted for human support without adequate safeguards. Where commissioning decisions prioritise cost reduction or efficiency over clinical appropriateness and ethical considerations, the resulting risks should be recognised as organisational responsibilities.

Individual clinicians should not be held liable for adverse outcomes arising from AI tools where they have acted within the scope of their role, training, and organisational guidance. Clinicians do not have visibility into how AI systems are built, trained, or updated, and should not be expected to identify hidden risks or biases embedded within them. Liability should only arise where there is clear evidence of professional misconduct or misuse that falls outside agreed protocols.

Service users should not bear liability for harms arising from AI technologies presented to them as safe, supportive, or therapeutic. Given the vulnerability that often accompanies mental distress, it would be inappropriate and unethical to assign responsibility to people who are engaging with systems in good faith, particularly where there may be limited understanding of how those systems operate or what their limitations are.

Question 9: Do you have any other evidence to contribute?

Please find the following documents attached along with this document.

- NCPS Relational Safeguards document
- NHS 10 Year Workforce Plan Submission – Section 1
- The importance of Human Connection and the impact of Digitalisation on Counselling Training and Practice
- Public Perceptions of AI and Counselling & Psychotherapy in Mental Health Support – May 2024
- Human Connection: Why it's vital in mental health support services