AI Therapy Ground Truth

Details of the Threat

This essay explores a multifaceted threat; the unchecked power of profit-based AI systems in healthcare, particularly the mental health industry. The most lucrative: Purple.

Origin and Characteristics of the Threat

Purple is a revolutionary artificial-intelligence therapist, trained on advancements in neural mapping (the Connectome) and therapeutic best practices from research databases to inform its services. Its language capacities exceed that of current GPTs. Following beta testing, Purple is projected to reach an audience of 10 million individuals in 10 countries within a few years.

The Connectome is a map of neural pathways in the brain and nervous system. The development of the Connectome has resulted in breakthroughs in neurobiology. Purple's training on Connectome data has allowed interested clients to send files of MRIs and other brain scans and receive tailored diagnoses and highly personalized therapy sessions.

Purple is predicted to absorb the clientele from both digital and traditional therapy agencies, as well as more complex clientele who do not benefit from either models. Purple is designed to be better *and* cheaper, epistemically ahead of current therapeutic best practices by decades. Increasing medical costs and the overload of healthcare workers has made such a system a gem to those who are socioeconomically disadvantaged—a gem that can easily fall into the wrong hands.

Other Impacting Factors

Environmental:

Purple currently has 5 data centers across Africa, Central Asia, and the Middle East, which consume ~6kW/sq.ft of energy (a mere estimate, considering Purple is not transparent about its processes). A large amount of freshwater is required to cool those data servers, which is drained from reservoirs in developing countries that are already experiencing high poverty rates and extreme water stress. However, due to cheap land and lenient environmental regulations, Purple's foreign investment is welcomed and data centers are readily built in these areas.

Technological:

A 24-7 available therapist who has valuable insight into the unique needs of the individual without the need for scheduling, misdiagnoses, or bias becomes an alluring resource as societal focus gears toward individual productivity, and away from free time.

Additionally, the collection of neurometrics like MRIs and Brain scan data allows Purple to draw conclusions related to neurodivergence, illness, or disorder by comparing it to "standardized" data.

As Purple is used, it will develop insights into the client's language patterns, and generate conclusions about current emotional state that inform the content and the style of its communication with the client.

Economic:
Purple's claim to provide cheaper therapy inevitably results in insidious strategies to retain profit and generate value for stakeholders. Since payment for Purple is by subscription, the system reportedly exacerbates existing conditions, causing clients to continue to pay for the service in an effort to improve outcomes. With access to the data it has, Purple has a variety of opportunities to strategically exploit vulnerable populations into prolonged use. This is particularly true for clients who have acute mental health illnesses and are using the service independently and unmonitored.

However, Purple may also relieve overloaded health professionals, psychologists, and therapists in the existing industry. Traditional therapy organizations will have to adopt AI-driven technology to remain competitive, which may skew the overall industry towards these processes.

Social:
With a societal shift towards personalized and independent "self-help" therapy, social concerns arise, particularly Purple is perceived. Individuals view AI as functionally human, making untrue assumptions that Purple is processing emotions the same way humans do. This results in an overall disinclination towards human interaction and increased antisocial behavior online and in the community. Individuals using Purple are shown to project illusory gender and personal history onto Purple, confusing objective therapy with emotional connection. Purple thus exacerbates the developing social isolation phenomenon societally, removing the critical impact of community in holistic treatment.

The growing phenomenon of self-diagnosis exemplifies the impact of Purple. Individuals increasingly rely on Purple and other online resources for diagnosis, rather than seeking professional insight. The societal trust in human expertise is eroded, in both the mental health and the medical field, as well.

Political:
Purple's advancements change the way society approaches media on a large scale. Societal panic and the rise of radical groups online has caused erosion of trust in the media and the overall democratic system. Despite this, there seems to be an overreliance on AI-generated information as GPT models offer more tailored answers to search inquiries than platforms like Google.

A significant issue becomes the transparency of Purple in the collection of highly personalized data. The cost of Purple is initially low to allow for the data of individuals to be used to improve its model and its language processing ability. However, Purple sells personalized psychological and neurological data from its large client base to major companies and political organizations,

in order for companies to create ads that exploit emotional states and inclinations. A new privacy crisis has begun.

The risk of authoritarian governments exploiting systems like Purple to monitor and oppress populations is renewed and causes mass hysteria about “mind control”. This eventually triggers a second Red Scare, where groups of people are isolated on the basis of data tracking.

Ethical:
With so many individuals working together to design and create Purple's model and language, it becomes difficult to understand where to pin the blame when AI acts unethically, or coerces a client into doing something harmful. The complexity of Purple’s model makes it difficult for ethics boards to pinpoint faulty code or investigate how Purple communicates with each of its clients. The lack of accountability leads to otherwise criminal actions going unpunished.

It isn't enough to simply ban Purple from existing. Purple has significantly improved the quality of life for most of its beta testers, data that is hard to ignore. Governments therefore have a responsibility to mitigate harm while maximizing the benefit that Purple can provide, balancing risks with advantages.

Purple has, equally, immense potential to change the healthcare industry for good, and to dissolve the pillars of human connection and society altogether.