

La ética de la IA en la psiquiatría

Centro de Investigación en Computación, IPN.

Miguel Angel Soto Hernandez.

En estos últimos años se han estado desarrollando aplicaciones para ayudar a personas que padecen enfermedades mentales. Estas aplicaciones pueden verse de distintas maneras como: emails, chatbots, aplicaciones móviles, etc. Y todas ellas tienen algo en común que es que hacen uso de modelos de inteligencia artificial como ayuda para los expertos ya sea psiquiatras o psicólogos al dar un diagnóstico de cierta manera más rápido. Sin embargo, de algún lugar se tiene que obtener información valiosa para que estos modelos les sean útiles a los expertos y para esto tenemos que obtener información ya sea de notas que tome un psiquiatra, de publicaciones en las distintas redes sociales del paciente o incluso que el mismo modelo aprenda constantemente de lo que se le dice para que mejore con el tiempo. Este es uno de los puntos más importantes ya que los modelos pueden llegar a su autonomía conforme más aprendan, pero ¿Cómo sabemos que nuestra información personal se encontrará a salvo?, ¿Qué pasa si el diagnóstico de la IA es erróneo?, Si la IA le hace daño al paciente ¿Quién se hace responsable?

A pesar de que estas implementaciones sean beneficiosas como en el caso de medicina preventiva como lo podría ser para el caso de prevenir el suicidio, por mencionar un ejemplo, y que las personas están dispuestas a platicar y llevar una conversación sobre problemas personales o en este caso problemas mentales con un ente artificial. No todas las personas lo ven de esta manera debido a que nos podemos hacer muchas preguntas con respecto al cómo funcionan o a que es lo que hacen con la información personal que nosotros al interactuar con ellos podemos darles. Es entonces donde entramos con la ética que pueden llegar a tener estos modelos de inteligencia artificial en el campo de la psiquiatría. Para esto hay que partir de que la ética no es simplemente un ejercicio de marcar casillas de que es lo que está bien y que es lo que está mal, ya que va más allá, y los principios deben ser complementados con virtudes morales.

Viéndolo de una perspectiva muy optimista, podríamos decir que no solo estaría procesando nuestra información, sino el de miles o millones de personas y si el modelo se encuentra bien estructurado no tendría porque tomar los nombres de los pacientes o información más personal para llevar a cabo su tarea, sino simplemente tomar la idea un poco más general con los posibles activadores que pueden desencadenar la enfermedad. Ahora, suponiendo que el manejo de datos personales no sea el problema, podríamos mencionar el hecho de los diagnósticos, que en este caso como somos muy optimistas, este casi nunca fallaría y podría eficientar el trabajo de los expertos.

Por otra parte, si vemos todo este asunto de las inteligencias artificiales de una manera más realista no podemos asegurar que el modelo siempre funcione como queremos que funcione ya que al final de cuentas estos modelos son una caja negra, y esto nos llevaría de vez en cuando a un mal diagnóstico y por ende que el paciente empeore en lugar de mejorar ya que no sabemos con total seguridad lo que hacen para procesar la información. Esto al contrario de un psiquiatra tiene muchas desventajas, ya que uno como persona al ir a terapia crea un vínculo doctor-paciente que no se puede romper. Agregándole a esto, el hecho de que una ente computacional llegue a una autonomía y falle porque en algún punto fallará y nadie asuma responsabilidad por x o y causaría mucha euforia por parte de las demás personas a tal grado que no confiarían de nuevo en una IA.