

## Transparencia y sus límites en los algoritmos responsables

Un valor muy importante para la rendición de cuentas en la toma de decisiones automatizada es la transparencia. Sin embargo, con respecto a los algoritmos que hacen grandes decisiones, como si se debe aprobar un préstamo o cuál será la regla para elegir algunas personas para ser auditadas por el departamento de impuestos sobre la renta, o cómo se asignan los fondos del gobierno, la transparencia podría no ser la mejor respuesta. Sin embargo, incluso cuando se divulga el código y se aclara la lógica de los algoritmos como una técnica para lograr que la práctica sea justa y no discriminatoria, de hecho, el software introduce una serie de problemas técnicos y éticos que dejan de hacerlo factible.

Primero, muchos de los algoritmos, en particular los que implementan enfoques de aprendizaje automático, son difíciles de la naturaleza misma. En estos casos, las reglas no están realmente programadas de manera explícita, sino que son emergentes, es decir, derivadas, de la estructura de los datos que el algoritmo está procesando. Esta propiedad, a la que se suele referir como "la caja negra", implica que ni siquiera al programador es fácil explicar cómo se ha llegado a un resultado particular. En tales circunstancias, la transparencia del código puede ser bastante no informativa, ya que el comportamiento del sistema está fuertemente influenciado no solo por los datos de entrada, sino también por la parametrización del modelo, que a menudo muchos luchan por entender.

Además, un exceso de hiper-transparencia puede ser contraproducente. Por ejemplo, en sistemas que determinan quién es auditado por el Servicio de Impuestos Internos o en la detección de proyectiles en la seguridad de un aeropuerto, un registro de los detalles de los criterios de decisión permitiría a los individuos hostiles aprender a eludir el sistema. Del mismo modo, en los sistemas empresariales, como los de toma de decisiones de crédito, la divulgación de todos los detalles podría divulgar información personal extremadamente privada o revelar secretos comerciales, poniendo en riesgo la privacidad y la competencia.

Este es el desafío que nos está obligando a maneras mucho más elegantes de pensar en la rendición de cuentas. En lugar de tratar de hacer plenamente transparentes los algoritmos, podríamos depender de mecanismos de prueba como auditorías independientes, pruebas de equidad y herramientas criptográficas a través de las cuales pudiéramos probar que se habían seguido las reglas sin hacer plenamente transparentes todos los detalles del algoritmo.