5

DISCUSIÓN

De estos hallazgos emergen dos tensiones transversales. La primera es la brecha principio–práctica: incluso donde las partes interesadas se ponen de acuerdo sobre principios de alto nivel, la traducción a mecanismos accionables de diseño, gobernanza y ejecución sigue siendo incompleta (Ashok et al., 2021; Morley et al., 2019). La segunda es la tensión entre armonización y pluralismo: el acuerdo global sobre una IA ética es deseable, pero los intentos de imponer estándares universales corren el riesgo de obliterar el pluralismo cultural y moral (Jobin et al., 2019), y las acusaciones de hegemonía cultural en la ética de la IA siguen vivas (Goffi, 2021). Los hallazgos implican también que los problemas más difíciles del campo no son puramente técnicos. Los debates sobre alineación, equidad y rendición de cuentas están todos subdeterminados por la ingeniería sola, lo que sugiere que el progreso depende de la colaboración interdisciplinaria entre tecnólogos, éticos, formuladores de políticas y el público (Hagendorff, 2020). Para la investigación, la agenda apunta a desarrollar métodos éticos sensibles a la situación en lugar de listas de verificación universales (Hagendorff, 2020); para la práctica, a incrustar la ética a lo largo de todo el ciclo de vida de la IA (Prem, 2023); y para la política, a mecanismos regulatorios adaptativos que puedan seguir el ritmo del cambio rápido sin sofocar la innovación (Dwivedi et al., 2019).

Las perspectivas críticas adicionales agudizan estas tensiones. El desafío de Kaczynski, leído como discurso y no como programa (Barnett, s.f.), funciona como una condición de frontera evaluable: la gobernanza, la alineación y la rendición de cuentas deben demostrarse en la práctica, y cuando los incentivos tienen la estructura de un dilema colectivo, saber que un riesgo existe

no basta para corregirlo (Falk & Tsoukalas, 2026; Moen, 2018). La respuesta no es la abolición, sino el diseño de instituciones e incentivos cuya eficacia pueda verificarse. La crítica epistemológica, a la inversa, constriñe al campo desde dentro: los sistemas generan implicaciones de representaciones elegidas, y cada elección —de lógica, de granularidad, de categoría, de valor— es una decisión humana que ningún progreso técnico puede convertir en un oráculo (Lee, 2024; Sowa, 1999). Juntas, las dos críticas niegan al campo sus ambiciones más fuertes desde direcciones opuestas, y ambas refuerzan la visión de que la ética de la IA no es un problema que deba resolverse una vez y definitivamente, sino una práctica que debe sostenerse.