4

RESULTADOS

La síntesis arroja seis hallazgos principales. Primero, la alineación de valores sigue sin resolverse tanto en el frente técnico como en el normativo: no hay un método acordado para codificar los valores humanos, ni un conjunto acordado de valores que codificar, ni una autoridad acordada que elegir entre ellos (Gabriel, 2020; Sutrop, 2020). Segundo, la investigación sobre equidad carece de una definición asentada de equidad, enfrenta múltiples fuentes de sesgo irreductibles y aún no ha producido una mitigación de extremo a extremo que abarque datos, modelo e implementación (Jui & Rivas, 2024; Pessach & Shmueli, 2022). Tercero, la explicabilidad y la rendición de cuentas están ligadas: los modelos opacos socavan tanto la justificación como la responsabilidad, y la existencia y naturaleza de la brecha de responsabilidad siguen siendo filosóficamente disputadas (Königs, 2022; Petersen & Zech, 2024; Tigard, 2020). Cuarto, el diseño de la moral de las máquinas está bloqueado por obstáculos empíricos y normativos a la vez —la heterogeneidad cultural de las preferencias morales y la ausencia de cualquier marco sobre el que el público y los reguladores puedan ponerse de acuerdo (Bonnefon et al., 2016; Lin, 2021). Quinto, la gobernanza va a la zaga del progreso tecnológico, con problemas abiertos en la operacionalización, la auditoría, el cumplimiento y la identificación de dimensiones regulatorias objetivo (Birkstedt et al., 2023; Reuel et al., 2024). Sexto, la viabilidad misma de una solución está en cuestión, dada la abstracción de los principios, la proliferación de herramientas sin conmensurabilidad y el fundamento normativo no resuelto de la propia empresa (Prem, 2023; Stahl et al., 2020).

Un séptimo hallazgo transversal concierne al uso cotidiano: la responsabilidad se distribuye entre diseñadores, tomadores de

decisiones, implementadores, auditores, usuarios, organizaciones, reguladores y organismos internacionales a lo largo del ciclo de vida de diseño–desarrollo–implementación (Liu et al., 2022; Novelli et al., 2023; Ryan & Stahl, 2020; Walter, 2024), pero el compartir mismo puede difuminar la responsabilidad, y si existe una brecha de responsabilidad genuina sigue siendo disputado (de Sio & Mecacci, 2021; Königs, 2022; Tigard, 2020). Finalmente, tres desafíos críticos acotan las ambiciones más fuertes del campo. La literatura sobre aprendizaje continuo muestra que el progreso aparente en el dilema estabilidad-plasticidad refleja a menudo una estabilidad comprada a expensas de la plasticidad (Kim & Han, 2023). La crítica radical contra la tecnología niega que el sistema tecnológico pueda reformarse desde dentro (Moen, 2018; Yampolskiy, 2022). Y la epistemología de la representación del conocimiento muestra que las afirmaciones de precisión y equidad universales son injustificadas: un sistema genera implicaciones de representaciones elegidas por humanos, y la solidez no puede ser certificada por ninguna maquinaria que razone sobre ellas (Lee, 2024; Servín, s.f.-a; Sowa, 1999).