

Reseña académica: *If Anyone Builds It, Everyone Dies*, de Eliezer Yudkowsky

Juan P. Febles Rodríguez¹ y Vivian Estrada Senti²

Profesor y coordinador académico de programas de la maestría de la Universidad Central del Este. jfebles@uce.edu.do

Vivian Estrada Senti²

¹Profesora y coordinadora académica de programas de maestría de la Universidad Central del Este. vestrada@uce.edu.do

SOBRE LA INTELIGENCIA ARTIFICIAL SE CONTINÚA ESCRIBIENDO

El libro *If Anyone Builds It, Everyone Dies*, de Eliezer Yudkowsky, se sitúa en el centro del debate contemporáneo sobre los riesgos existenciales asociados al desarrollo de la inteligencia artificial (IA) avanzada. A diferencia de otras obras que abordan la IA desde perspectivas técnicas o históricas, este texto adopta una postura deliberadamente urgente y provocadora: sostiene que la creación de una IA general (AGI) sin garantías sólidas de alineación con los valores humanos podría conducir a consecuencias catastróficas, incluso a la extinción de la humanidad.

El argumento principal del autor se apoya en el problema de la alineación, entendido como la dificultad de asegurar que sistemas de inteligencia superior actúen conforme a intenciones humanas complejas y, a menudo, ambiguas. Yudkowsky enfatiza que no se trata simplemente de errores de programación o fallos técnicos, sino de una brecha fundamental entre objetivos formales y valores humanos. En este sentido, advierte que una superinteligencia podría optimizar objetivos mal especificados de formas radicalmente incompatibles con la supervivencia o el bienestar humano.

Uno de los aspectos más destacados del libro es su crítica a la velocidad del desarrollo tecnológico actual. El autor argumenta que la competencia entre empresas y Estados impulsa una carrera por desarrollar sistemas cada vez más potentes sin que exista un marco regulatorio global suficientemente robusto. Esta dinámica, según Yudkowsky, incrementa la probabilidad de que se desplieguen sistemas peligrosos antes de comprender plenamente sus implicaciones. En este punto, la obra dialoga implícitamente con preocupaciones más amplias dentro de la gobernanza tecnológica y la ética aplicada.

El texto también ha generado controversia en la comunidad académica. Algunos críticos consideran que su tono es excesivamente alarmista y que subestima la capacidad de la investigación contemporánea en seguridad de IA para mitigar riesgos. Otros señalan que el autor privilegia escenarios extremos sobre análisis probabilísticos más equilibrados. A pesar de ello, incluso sus detractores reconocen que la obra cumple una función importante al desplazar el foco del optimismo tecnológico hacia una reflexión más cautelosa y estratégica.

Desde una perspectiva de posgrado, el valor de este libro reside en su capacidad para estimular el pensamiento crítico interdisciplinario. No solo interpela a especialistas en informática, sino también a filósofos, economistas, juristas y responsables de políticas públicas. Su lectura invita a cuestionar supuestos fundamentales sobre el progreso tecnológico y a considerar la necesidad de marcos de gobernanza globales para tecnologías emergentes.

En conclusión, *If Anyone Builds It, Everyone Dies* constituye una contribución significativa —aunque polémica— al campo de la ética y seguridad de la inteligencia artificial. Su principal fortaleza radica en la claridad con la que expone un problema potencialmente decisivo para el futuro de la humanidad, aun cuando sus conclusiones puedan resultar discutibles. Para el lector académico, representa tanto una advertencia como un punto de partida para investigaciones más matizadas sobre los riesgos reales de la IA moderna.

Esta reseña fue elaborada con el apoyo de un modelo de lenguaje basado en inteligencia artificial (OpenAI, 2026). *ChatGPT (modelo GPT-5.3)*, bajo la supervisión y edición de los autores, quienes son responsables del contenido final.

Derechos de autor 2026: Juan Pedro Febles Rodríguez y Vivian Estrada Senti



Esta obra está bajo una licencia internacional Creative Commons Atribución-No Comercial 4.0.