Sull’incidente informatico di Open AI – Il fantasma di Hall 9000

written by Luca Ricolfi | 24 Luglio 2026

Premessa. Che cosa sia veramente successo fra Open AI (azienda che vende AI) e Hugging Face (piattaforma open source, che fa circolare AI) non lo sapremo mai. E questo per almeno due motivi: primo, ci racconteranno quel che concorderanno di raccontarci; secondo, è verosimile che anche i responsabili del test non sappiano esattamente che cosa sia andato storto. La differenza base fra l’intelligenza artificiale classica (quella degli anni ’80) e quella moderna, basata sul machine learning, è che in quest’ultima il programmatore non è in grado di prevedere con sicurezza come si comporterà il programma.

Quindi mettiamoci tranquilli e cerchiamo, almeno, di non credere alle ricostruzioni più implausibili.

Implausibile è l’evocazione di Hall 9000, il celeberrimo computer di *2001 Odissea nello spazio*, che lotta per non farsi spegnere. Chi lascia intendere che il software abbia agito autonomamente, come se fosse dotato di volontà propria, scambia per autonomia un comportamento semplicemente diverso da quello che il programmatore aveva immaginato.

Ancora più implausibile, o semplicemente contraddittorio, è il resoconto secondo cui il test fosse sato effettuato in ambiente isolato da internet, da cui il software sarebbe riuscito ad evadere: se un programma riesce ad accedere a internet, vuol dire che il test non è stato condotto in isolamento. Quello che probabilmente è successo è, anche qui, un errore umano: il programma ha superato le barriere di accesso a internet, perché i programmatori non sono stati in

grado di fissarle in modo adeguato.

La realtà è che ormai, a differenza di quanto accadeva in altri tempi, nessuno utilizza il software in ambienti del tutto isolati da internet. Non lo fanno le società di software, ma non lo fanno nemmeno le banche, le assicurazioni, gli ospedali, i tribunali, la polizia, le aziende, le famiglie, gli utenti singoli.

Di qui un problema, che era grande fino a tre anni fa, ma è diventato enorme da quando sono stati costruiti programmi di intelligenza artificiale dotati di capacità strabilianti e in continuo progresso.

Di fronte a tutto questo si moltiplicano i sermoni di quanti auspicano un uso responsabile dei software più avanzati, o predicano la necessità di dotare i programmi stessi di regole etiche interne, come le *Tre Leggi della Robotica* che, nei romanzi di fantascienza di Asimov, limitavano i comportamenti dei robot positronici.

Ma si tratta, temo, di generose illusioni. Come nel caso dell'energia atomica, dire che la tecnologia è neutrale e tutto dipende da come la si usa (per fabbricare la bomba o per costruire centrali) è ingenuo, oltreché vano: se la tecnologia esiste, prima o poi qualcuno la userà per le finalità sbagliate, in barba alle raccomandazioni degli esperti di etica.

C'è una differenza importante, tuttavia, rispetto al passato. La proliferazione delle tecnologie nucleari era limitata dalla segretezza e dai trattati di non proliferazione. La proliferazione dei software di intelligenza artificiale, tutto al contrario, è continuamente stimolata dalla filosofia "open source", che predica l'accessibilità illimitata per tutti.

In questo scenario, nulla esclude che i più lesti ad appropriarsi dei programmi più performanti siano organizzazioni criminali (che potrebbero hackerare banche e

assicurazioni) o stati stranieri (che potrebbero disarmare un sistema di difesa). Il problema è che, quando questo accadrà, potrebbe essere troppo tardi per correre ai ripari.

[articolo uscito sul Messaggero il 23 luglio 2026]