Capitolo primo
Se ci riusciamo

Molto tempo fa i miei genitori vivevano a Birmingham, in Inghilterra, in una casa vicino all'università. Decisero di andarsene dalla città e vendettero la casa a David Lodge, un professore di letteratura inglese. All'epoca Lodge era già un famoso romanziere. Non l'ho mai conosciuto ma ho deciso di leggere alcuni suoi libri: *Scambi* e *Il professore va al congresso*. Tra i personaggi principali c'erano accademici di fantasia che si trasferivano da una Birmingham di fantasia a una Berkeley, California, di fantasia. Poiché io ero un vero accademico della vera Birmingham che si era appena trasferito nella vera Berkeley, sembrava che qualcuno del Dipartimento delle Coincidenze mi stesse invitando a prestare attenzione.

Una scena in particolare de *Il professore va al congresso* mi ha colpito: il protagonista, aspirante teorico della letteratura, partecipa a un'importante conferenza internazionale e chiede a una tavola rotonda di figure illustri: «Che cosa succede se tutti sono concordi con voi[?]»[1]. La domanda provoca un certo sgomento, perché ai partecipanti interessa di piú lo scontro intellettuale che accertare la verità o giungere a un'intesa. Mi è venuto in mente che si potrebbe rivolgere una domanda simile alle figure piú illustri dell'IA: «E se ci riuscite?» L'obiettivo del settore è sempre stato quello di creare un'IA a livello umano o sovrumano, ma si è riflettuto poco o niente su cosa succederebbe se lo facessimo.

Qualche anno piú tardi, io e Peter Norvig cominciammo a lavorare a un nuovo manuale di IA, la cui prima edizione

uscí nel 1995[2]. L'ultima sezione del libro è intitolata «Che succede se ci riusciamo?» e descrive i possibili esiti positivi e negativi, pur senza giungere a conclusioni definitive. Quando nel 2010 è uscita la terza edizione, finalmente molti hanno cominciato a prendere in considerazione la possibilità che l'IA sovrumana non sia una buona cosa, ma si trattava soprattutto di outsider, non dei principali ricercatori nel campo dell'IA. Nel 2013 ero ormai convinto che la questione non solo era entrata nel discorso popolare ma che probabilmente si trattava del problema piú importante che l'umanità avesse davanti a sé.

Nel novembre del 2013 ho tenuto una conferenza alla Dulwich Picture Gallery, un venerabile museo nella zona sud di Londra. Il pubblico era formato per lo piú da pensionati, persone estranee alla scienza che nutrivano un interesse generico per le questioni intellettuali, cosí ho dovuto tenere un discorso che esulava completamente dagli aspetti tecnici. Sembrava il posto adeguato per testare le mie idee in pubblico per la prima volta. Dopo aver spiegato cos'è l'IA, ho nominato cinque dei possibili «principali eventi del futuro dell'umanità»:

1. Moriamo tutti (per l'impatto di un asteroide, una catastrofe climatica, una pandemia ecc.).
2. Viviamo tutti per sempre (per una soluzione medica all'invecchiamento).
3. Inventiamo un modo per viaggiare piú velocemente della luce e conquistiamo l'universo.
4. Riceviamo la visita di una civiltà aliena superiore.
5. Inventiamo un'IA superintelligente.

Ho affermato che la quinta ipotesi, quella dell'IA superintelligente, avrebbe vinto perché ci avrebbe aiutato a evitare catastrofi fisiche e a ottenere la vita eterna e i viaggi piú veloci della luce, se questi fossero stati possibili. Avrebbe rappresentato un grande salto – una discontinuità – per la nostra civiltà. L'avvento dell'IA superintelligente è per molti versi

analogo all'arrivo di una civiltà aliena superiore ma è molto piú probabile. Forse il dato piú importante è che l'IA, a differenza degli alieni, è una cosa su cui abbiamo voce in capitolo.

Poi ho chiesto al pubblico di immaginare cosa succederebbe se una civiltà aliena superiore ci avvisasse che sta per arrivare sulla Terra tra trenta-cinquant'anni. La parola *caos* non rende neanche minimamente l'idea. Tuttavia la nostra reazione all'arrivo previsto dell'IA superintelligente è stata… be', deludente rende bene l'idea. (In una conferenza successiva, ho illustrato quest'ipotesi sotto forma di uno scambio di e-mail, alla figura 1). Infine ho spiegato l'importanza dell'IA superintelligente come segue: «Riuscirci sarebbe l'evento piú grosso della storia umana… e forse l'ultimo evento della storia umana».