Priorità di ricerca per un'intelligenza artificiale robusta e benefica: una lettera aperta

Vi è ora un ampio consenso sul fatto che la ricerca sull'IA stia progredendo costantemente e che il suo impatto sulla società sia destinato ad aumentare. I potenziali benefici sono enormi, poiché tutto ciò che la civiltà ha da offrire è un prodotto dell'intelligenza umana. A causa del grande potenziale dell'IA, è importante ricercare come trarne i benefici evitando potenziali insidie.

28 ottobre 2015

La ricerca sull'intelligenza artificiale (AI) ha esplorato una varietà di problemi e approcci sin dal suo inizio, ma negli ultimi 20 anni circa si è concentrata sui problemi che circondano la costruzione di agenti intelligenti, sistemi che percepiscono e agiscono in un determinato ambiente. In questo contesto, "intelligenza" è correlata alle nozioni statistiche ed economiche di razionalità - colloquialmente, la capacità di prendere buone decisioni, piani o inferenze. L'adozione di rappresentazioni probabilistiche e teoriche della decisione e metodi di apprendimento statistico ha portato a un ampio grado di integrazione e fertilizzazione incrociata tra intelligenza artificiale, apprendimento automatico, statistica, teoria del controllo, neuroscienze e altri campi. La definizione di quadri teorici condivisi, unita alla disponibilità di dati e potenza di elaborazione,
Poiché le capacità in queste e altre aree superano la soglia dalla ricerca di laboratorio alle tecnologie economicamente valide, si instaura un circolo virtuoso per cui anche piccoli miglioramenti nelle prestazioni valgono ingenti somme di denaro, spingendo maggiori investimenti nella ricerca. Vi è ora un ampio consenso sul fatto che la ricerca sull'IA stia progredendo costantemente e che il suo impatto sulla società sia destinato ad aumentare. I potenziali benefici sono enormi, poiché tutto ciò che la civiltà ha da offrire è un prodotto dell'intelligenza umana; non possiamo prevedere cosa potremmo ottenere quando questa intelligenza viene amplificata dagli strumenti che l'intelligenza artificiale può fornire, ma l'eradicazione delle malattie e della povertà non è insondabile. A causa del grande potenziale dell'IA, è importante ricercare come trarne i benefici evitando potenziali insidie.
I progressi nella ricerca sull'IA rendono opportuno concentrare la ricerca non solo sul rendere l'IA più capace, ma anche sulla massimizzazione dei benefici sociali dell'IA. Tali considerazioni hanno motivato il Panel presidenziale AAAI 2008-09 sui futuri dell'IA a lungo termine e altri progetti sugli impatti dell'IA, e costituiscono una significativa espansione del campo dell'IA stesso, che fino ad ora si è concentrato in gran parte su tecniche neutre rispetto a scopo. Raccomandiamo una ricerca ampliata volta a garantire che i sistemi di intelligenza artificiale sempre più capaci siano robusti e vantaggiosi: i nostri sistemi di intelligenza artificiale devono fare ciò che vogliamo che facciano. In allegato il documento sulle priorità di ricercafornisce molti esempi di tali direzioni di ricerca che possono aiutare a massimizzare il vantaggio sociale dell'IA. Questa ricerca è necessariamente interdisciplinare, perché coinvolge sia la società che l'intelligenza artificiale. Si spazia

dall'economia, diritto e filosofia alla sicurezza informatica, metodi formali e, naturalmente, vari rami dell'IA stessa.
In sintesi, riteniamo che la ricerca su come rendere i sistemi di intelligenza artificiale robusti e vantaggiosi sia importante e opportuna e che vi siano indicazioni di ricerca concrete che possono essere perseguite oggi.
In caso di domande su questa lettera, contattare Max Tegmark .

Aggiungi la tua firma

Firmatari

Fare clic qui per visualizzare l'elenco completo dei firmatari.
Ad oggi, la lettera aperta è stata firmata da oltre 8.000 persone. L'elenco dei firmatari comprende:

Firmatari di spicco

Stuart Russell , Berkeley, professore di informatica, direttore del Center for Intelligent Systems e coautore del libro di testo standard Artificial Intelligence: a Modern Approach.
Tom Dietterich , Stato dell'Oregon, Presidente di AAAI, Professore e Direttore dei Sistemi Intelligenti
Eric Horvitz , direttore della ricerca Microsoft, ex presidente dell'AAAI, co-presidente del panel presidenziale dell'AAAI sui futuri dell'IA a lungo termine
Bart Selman , Cornell, professore di informatica, co-presidente del panel presidenziale AAAI sui futuri dell'IA a lungo termine
Francesca Rossi , Padova e Harvard, Professore di Informatica, Presidente IJCAI e Copresidente del comitato AAAI sull'impatto dell'IA e questioni etiche
Demis Hassabis , co-fondatrice di DeepMind
Shane Legg , co-fondatore di DeepMind
Mustafa Suleyman , co-fondatore di DeepMind
Dileep George , co-fondatore di Vicarious
Scott Phoenix , co-fondatore di Vicarious
Yann LeCun , capo del laboratorio di intelligenza artificiale di Facebook
Geoffrey Hinton , Università di Toronto e Google Inc.
Yoshua Bengio , Université de Montréal
Peter Norvig , direttore della ricerca presso Google e coautore del libro di testo standard Artificial Intelligence: a Modern Approach
Oren Etzioni , CEO di Allen Inst. per l'IA
Guruduth Banavar , VP, Calcolo cognitivo, IBM Research
Michael Wooldridge , Oxford, capo del dipartimento di informatica, presidente del comitato di coordinamento europeo per l'intelligenza artificiale
Leslie Pack Kaelbling , MIT, professoressa di informatica e ingegneria, fondatrice del Journal of Machine Learning Research
Tom Mitchell , CMU, ex presidente di AAAI, presidente del dipartimento di Machine Learning
Toby Walsh , Univ. del New South Wales e NICTA, Professore di IA e Presidente della AI Access Foundation
Murray Shanahan , Imperial College, professore di robotica cognitiva
Michael Osborne , Oxford, Professore Associato di Machine Learning

David Parkes , Harvard, professore di informatica
Laurent Orseau , Google DeepMind
Ilya Sutskever , Google, ricercatore di intelligenza artificiale
Blaise Aguera y Arcas , Google, ricercatore di intelligenza artificiale
Joscha Bach , MIT, ricercatore di intelligenza artificiale
Bill Hibbard , Madison, ricercatore di intelligenza artificiale
Steve Omohundro , ricercatore di intelligenza artificiale
Ben Goertzel , Fondazione OpenCog
Richard Mallah , Cambridge Semantics, direttore di Advanced Analytics, ricercatore di intelligenza artificiale
Alexander Wissner-Gross , Harvard, Fellow presso l'Institute for Applied Computational Science
Adrian Weller , Cambridge, ricercatore di intelligenza artificiale
Jacob Steinhardt , Stanford, AI Ph.D. alunno
Nick Hay , Berkeley, Ph.D. alunno
Jaan Tallinn , co-fondatore di Skype, CSER e FLI
Elon Musk , SpaceX, Tesla Motors
Steve Wozniak , co-fondatore di Apple
Luke Nosek , Fondo dei fondatori
Aaron VanDevender , Fondo dei fondatori
Erik Brynjolfsson , MIT, professore e direttore dell'iniziativa del MIT sull'economia digitale
Margaret Boden , U. Sussex, Professore di Scienze Cognitive
Martin Rees , Cambridge, professore emerito di cosmologia e astrofisica, vincitore di Gruber & Crafoord
Huw Price , Cambridge, Bertrand Russell Professore di Filosofia
Nick Bostrom , Oxford, professore di filosofia, direttore del Future of Humanity Institute (Oxford Martin School)
Stephen Hawking , Direttore della ricerca presso il Dipartimento di Matematica Applicata e Fisica Teorica di Cambridge, vincitore del Fundamental Physics Prize 2012 per il suo lavoro sulla gravità quantistica
Luke Muehlhauser , direttore esecutivo del Machine Intelligence Research Institute (MIRI)
Eliezer Yudkowsky , ricercatore del MIRI, co-fondatore del MIRI (allora conosciuto come SIAI)
Katja Grace , ricercatrice del MIRI
Benja Fallenstein , ricercatore del MIRI
Nate Soares , ricercatore del MIRI
Paul Christiano , Berkeley, studente laureato in Informatica
Anders Sandberg , Oxford, ricercatore del Future of Humanity Institute (Oxford Martin School)
Daniel Dewey , Oxford, ricercatore del Future of Humanity Institute (Oxford Martin School)
Stuart Armstrong , Oxford, ricercatore del Future of Humanity Institute (Oxford Martin School)
Toby Ord , Oxford, ricercatore del Future of Humanity Institute (Oxford Martin School), fondatore di Giving What We Can
Neil Jacobstein , Università della Singolarità
Dominik Grewe , Google DeepMind
Roman V. Yampolskiy , Università di Louisville
Vincent C. Müller , ACT/Anatolia College
Amnon H. Eden , Università dell'Essex
Henry Kautz , Università di Rochester

Boris Debic , Google, Chief History Officer
Kevin Leyton-Brown , Università della British Columbia, Professore di Informatica
Trevor Back , Google DeepMind
Moshe Vardi , Rice University, redattore capo delle comunicazioni dell'ACM
Peter Sincak , prof. TU Kosice, Slovacchia
Tom Schaul , Google DeepMind
Grady Booch , collega IBM
Alan Mackworth , Professore di Informatica, Università della British Columbia. Ex Presidente dell'AAA
Andrew Davison , Professore di Robot Vision, Direttore del Dyson Robotics Lab presso l'Imperial College di Londra
Daniel Weld , WRF / TJ Cable Professore di Informatica e Ingegneria, Università di Washington
Michael Witbrock , Cycorp Inc e AI4Good.org
Stephen L. Reed , ai-coin.com
Thomas Stone , co-fondatore di PredictionIO
Dan Roth , Università dell'Illinois, caporedattore di The Journal of AI Research (JAIR)
Babak Hodjat , Tecnologie senzienti
Vincent Vanhoucke , Google, ricercatore di intelligenza artificiale
Itamar Arel , Stanford University, professore di informatica
Ramon Lopez de Mantaras , direttore dell'Istituto di ricerca sull'intelligenza artificiale, Consiglio nazionale delle ricerche spagnolo
Antoine Blondau , Tecnologie senzienti
George Dvorsky , redattore collaboratore, io9; Presidente del consiglio di amministrazione dell'Istituto per l'etica e le tecnologie emergenti
George Church , Harvard e MIT
Klaus-Dieter Althoff , Università di Hildesheim, Professore di Intelligenza Artificiale; Responsabile del Competence Center Case-Based Reasoning, German Research Center for Artificial Intelligence, Kaiserslautern; Redattore capo della rivista tedesca sull'intelligenza artificiale
Christopher Bishop , Distinguished Scientist, Microsoft Research
Jen-Hsun Huang , CEO di NVIDIA