

di Roberto Rosso

I primi due giorni di novembre si è svolto a Londra l'AI Safety Summit¹ sull'Intelligenza Artificiale a cui hanno partecipato 28 paesi e l'Unione Europea. Il summit nasce dalla volontà del governo inglese di porre il proprio paese come polo a livello mondiale nel governo e lo sviluppo dell'ecosistema tecnologico dell'I.A.

Di per sé questo incontro fornisce la dimostrazione di come l'I.A. giochi sin da oggi -e sempre di più in futuro- un ruolo cruciale nella trasformazione dei rapporti economici e sociali a livello globale, nel governo delle società e nei rapporti strategici. Possiamo presumere che dietro le motivazioni che hanno spinto tanti governi a partecipare ci sia la consapevolezza di vivere in un contesto caratterizzato innanzitutto dall'avanzare del cambiamento climatico indotto dal riscaldamento globale e da ogni sorta di squilibri che vengono da un po' di tempo definiti come policrisi, intreccio di diversi processi di crisi che agiscono contemporaneamente ed in modo sinergico, rendendo sostanzialmente imprevedibili lo sviluppo degli eventi a livello regionale e globale. In questo contesto, nel quale l'orizzonte di imprevedibilità degli eventi è sempre più ravvicinato, diventa essenziale avere a disposizione strumenti in grado di interpretare, modellizzare, prevedere l'andamento di processi di complessità crescente.

Il summit ha prodotto una dichiarazione *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023*². Il carattere dell'incontro ed i tempi della sua preparazione non potevano produrre un testo programmatico ricco di contenuti strategici, comunque se fosse una dichiarazione sul cambiamento climatico lo potremmo collocare nel campo del greenwashing.

Il documento rilancia la cooperazione internazionale nel campo dell'I.A. e -come ogni analisi corrente- ne riconosce le straordinarie potenzialità trasformative ed indica nel contempo il coesistere di altrettanto straordinari rischi ed opportunità. *Si indica una direttrice per lo sviluppo e l'utilizzo dell'I.A. umano-centrica, affidabile e responsabile.* Se l'elenco delle opportunità definisce il quadro del migliore dei mondi possibili, proprio per il carattere pervasivo dell'utilizzo dell'I.A. nelle sue diverse forme, l'elenco dei rischi cita in buona sostanza la possibilità di sovvertire i dispositivi, le procedure con cui si può certificare l'attendibilità di una qualsiasi informazione, la possibilità di penetrare entro qualsiasi relazione e comportamento attraverso le sue tracce digitali e l'utilizzo delle stesse a fini genericamente criminali, in questo elenco la guerra non è citata.

La preoccupazione maggiore in realtà è orientata verso la frontiera degli sviluppi possibili dell'I.A., in questo condividendo l'orientamento prevalente nel confronto pubblico sul tema, l'attenzione è focalizzata su "highly capable general-purpose AI models, including foundation models, that could perform a wide variety of tasks - as well as relevant specific narrow AI that could exhibit capabilities that cause harm - which match or exceed the capabilities present in today's most advanced models" Centrale è il riferimento ai 'Foundation Models' a cui appartengono tutte le applicazioni di cui si parla negli ultimi mesi l'origine del termine è la seguente³ "The Stanford Institute for Human-Centered Artificial Intelligence's (HAI) Center for Research on Foundation Models (CRFM) coined the term "foundation model" in August 2021, tentatively referring to "any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks".[14] This was based on their observation that existing overlapping

terms were not adequate, submitting that “(large) language model’ was too narrow given [the]focus is not only language; ‘self-supervised model’ was too specific to the training objective; and ‘pretrained model’ suggested that the noteworthy action all happened after ‘pretraining.”[15] After considering many terms, they settled on “foundation model” to emphasize the intended function (i.e., amenability to subsequent further development) rather than modality, architecture, or implementation.”⁴

In merito alla definizione e ai rischi insiti nei cosiddetti foundation models quindi vale la pena di leggere il seguente documento stilato da un largo numero di esperti “On the Opportunities and Risks of Foundation Models”⁵ del Center for Research on Foundation Models (CRFM) Stanford Institute for Human-Centered Artificial Intelligence (HAI) Stanford University.

Il documento messo dal summit ha un focus sui futuri sviluppi della tecnologia, che sono in corso di elaborazione, di cui si hanno consistenti indizi, su cui ci si interroga senza poterne delineare le traiettorie effettive. Su questo andremo a dare qualche approfondimento più avanti.

L’obiettivo indicato dopo aver delineato rischi ed opportunità è quello di incrementare a partire dal summit -che si prevede di replicare nel 2024- forme di cooperazione tra tutti gli attori coinvolti nello sviluppo, utilizzo e controllo delle tecnologie di A.I.⁶

Si propone anche di creare una rete internazionale di centri di ricerca.⁷ dove l’enfasi è portata sulla sicurezza relativa alla nuova frontiera dell’I.A.

Il problema fondamentale è ritrovare un punto di equilibrio, posto che esista, tra governo del processo di innovazione -attraverso la individuazione e controllo dei rischi individuati- e lo stimolo dell’innovazione stessa.⁸ La definizione di classi , profili i rischio per le tecnologie di I.A. costituisce un fattore centrale per mettere in atto una qualsiasi sorta di governo dell’ecosistema dell’I.A., non a caso uno dei punti nodali dello EU A.I. Act la direttiva in corso di approvazione che regolerà tutte le attività di produzione e utilizzo della A.I. in Europa⁹.

Vedremo cosa si concretizzerà dei propositi usciti dal summit, vero è che lo sviluppo delle tecnologie di I.A. è uno strumento della competizione a livello globale, compreso l’uso duale cioè civile e militare, mentre le forme di governo sono poi connesse al carattere dei regimi politici di ogni singolo paese, vedi le differenze più volte evidenziate tra Cina, Stati Uniti ed Europa.

In proposito quasi in contemporanea al Summit di Londra il 30 ottobre è stato emesso dalla presidenza degli Stati Uniti un ordine esecutivo che definisce in modo molto puntuale e articolato la strategia del governo statunitense nei confronti dell’intero ecosistema tecnologico -ma anche economico, giuridico e sociale- dell’I.A. “President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence”¹⁰

L’elenco degli standard proposti dall’Executive Order (E.O.), sintetizzato nel Fact Sheet, è impressionante e costituisce una sintetica descrizione dei rischi individuati nello sviluppo ed utilizzo dell’I.A.; in precedenza l’amministrazione USA ha messo il Blueprint for an AI Bill of Rights a cui fa riferimento l’E.O. Il contesto in cui si colloca l’E.O. è definito da un altro E.O. del 2021 l’Executive Order to Strengthen Racial Equity and Support for Underserved Communities Across the Federal Government¹¹ a cui l’E.O. sulla I.A. fa riferimento in diversi punti tra i quali i seguenti.

Addresses emerging civil rights risks. The Executive Order instructs agencies to focus their civil rights authorities and offices on emerging threats, such as *algorithmic discrimination in automated*

technology; improve accessibility for people with disabilities; improve language access services; and consider opportunities to bolster the capacity of their civil rights offices. It further directs agencies to ensure that their own use of artificial intelligence and automated systems also advances equity. Promotes data equity and transparency. The Interagency Working Group on Equitable Data created by the day one Executive Order has been institutionalized at the National Science and Technology Council. This Executive Order directs the body to facilitate better collection, analysis, and use of demographic data to advance equity, and to regularly report on progress to the White House and the American public.

Questi due paragrafi evidenziano il carattere pervasivo del digitale nella società, la centralità della raccolta e condivisione di dati e informazioni, la spinta verso l'uso dell'I.A. in tutti gli apparati amministrativi e di governo.

Nell' E.O. sull'A.I. è illuminante l'elenco dei capitoli: New Standards for AI Safety and Security, Protecting Americans' Privacy, Advancing Equity and Civil Rights, Standing Up for Consumers, Patients, and Students, Supporting Workers, Promoting Innovation and Competition, Advancing American Leadership Abroad, Ensuring Responsible and Effective Government Use of AI.

Il primo paragrafo sugli standard di sicurezza recita.

Esigere che gli sviluppatori dei sistemi di IA più potenti condividano le loro risultati dei test di sicurezza e altre informazioni critiche con gli Stati Uniti governo. In conformità con il Defense Production Act, l'Ordine richiederà che le aziende che sviluppano un Foundation Model che ponga un grave rischio per la sicurezza nazionale, la sicurezza economica nazionale o la salute e la sicurezza pubblica debbano informare il Governo Federale quando addestrare il modello e condividere i risultati di tutte le operazioni di sicurezza del Test (*red team safety tests*).

L'attività di controllo e di sviluppo di dispositivi adatti allo scopo è affidata al National Institute of Standards and Technology - per quanto riguarda la definizione degli standard, il Department of Homeland Security dovrà applicare gli standard ai settori delle infrastrutture critiche ed istituire l'AI Safety and Security Board. I Departments of Energy and Homeland Security dovranno porre attenzione alle minacce che il sistema di I.A. possono costituire per le infrastrutture critiche come quelle di tipo chimico, biologico, radiologica ed ai rischi relativi alla cybersecurity. *Il connubio tra I.A. e biotecnologie è particolarmente sottolineato*, quando queste hanno obiettivi malevoli o criminali.

Per sovrintendere all'uso dell'I.A. in campo militare "Ordina lo sviluppo di un National Security Memorandum che diriga ulteriori azioni in materia di I.A. e sicurezza, che devono essere sviluppate dal National Security Council e dal White House Chief of Staff. Questo documento garantirà che l'esercito e la comunità dell'intelligence degli Stati Uniti utilizzi l'I.A. in modo sicuro, etico ed efficace nelle proprie missioni e dirigerà le azioni per contrastare l'uso militare dell'IA da parte degli avversari."

Il problema della discriminazione algoritmica già emersa in numerose applicazioni è ovviamente sottolineato mentre sul tema delicato della giustizia:

"Garantire l'equità in tutto il sistema di giustizia penale sviluppando le migliori pratiche sull'uso dell'IA nelle sentenze, nella libertà condizionale e nella libertà vigilata, liberazione preventiva e detenzione, valutazione dei rischi, sorveglianza, previsione dei crimini, polizia predittiva e analisi forense."

Se non siamo al livello del film *Minority Report*, diretto da Steven Spielberg, liberamente tratto dall'omonimo racconto di fantascienza di Philip K. Dick *Rapporto di minoranza*, è del tutto evidente come si preveda il dilagare delle procedure basate sull'I.A. in tutte le attività cosiddette di prevenzione del crimine e nella pratica giudiziaria in tutte le sue articolazioni; è difficile non cedere alla tentazione di profilare individui e comunità, automatizzando tutta una serie di procedure. Aggiungiamo che la problematica va ben oltre il campo che attiene in vario modo alle 'attività criminali' e riguarda l'analisi, la classificazione ed il governo della composizione sociale. L'impatto sulla organizzazione del lavoro e sulla condizione dei lavoratori è una problematica attuale, secondo ogni previsione destinata indubbiamente ad ampliarsi ed è già entrata nelle rivendicazioni e nei conflitti sindacali, in seguito alla sostituzione di funzioni lavorative con dispositivi di I.A.

"Per mitigare questi rischi, sostenere la capacità dei lavoratori di contrattare collettivamente e investire nella formazione della forza lavoro e uno sviluppo accessibile a tutti, il Presidente dirige le seguenti azioni."

Ciò che segue sono ovviamente azioni palliative a fronte di una riorganizzazione progressiva e radicale dell'organizzazione del lavoro in tutti i settori.

La parte finale si preoccupa di garantire il migliore ambiente per lo sviluppo dell'ecosistema tecnologico dell'I.A. supportando la nascita di nuovi soggetti dell'innovazione¹².

In buona sostanza si tratta per l'amministrazione di garantire il primato degli Stati Uniti nel campo delle tecnologie di I.A. il che significa -dato il loro carattere pervasivo e dirompente- mantenere il primato in ogni campo dell'innovazione tecnologica, fondamentale per vincere la competizione a livello globale, in particolare con la Cina, in un periodo in cui il primato indiscusso degli USA a livello globale è sostanzialmente venuto meno.

Questo Executive Order manifesta la piena consapevolezza da parte dell'amministrazione Biden del carattere dirompente delle tecnologie di I.A. -cuore e cervello dell'ecosistema digitale e dell'innovazione delle diverse filiere tecnologiche. Per questo suo carattere il suo sviluppo non può essere semplicemente lasciato alle dinamiche di mercato poiché può mettere in discussione caratteri fondamentali del sistema economico e sociale, anzi non v'è dubbio che le trasformazioni indotte in tempi più o meno lunghi saranno radicali e già cominciano ad operare; l'I.A. interviene sui meccanismi più profondi della riproduzione sociale e sugli strumenti di riproduzione del potere di fatto e degli equilibri istituzionali. Lo scenario in cui si pensa di operare prescinde dalle speculazioni più azzardate sul carattere di ciò che chiamiamo Intelligenza Artificiale, termine coniato in tempi in cui nemmeno si sospettava dell'esistenza dei dispositivi e delle tecnologie sviluppate negli ultimi anni. Più che focalizzarci semplicemente -si fa per dire- sulla struttura e le prestazioni di questi dispositivi è necessario più concretamente coglierne l'operatività nei diversi contesti, le loro prestazioni sociali e non solo di laboratorio. ChatGPT e gli altri Large Language Modules sono stati definiti 'pappagalli stocastici' in quanto la loro espressione è il prodotto della analisi, della rielaborazione in termini probabilistici delle relazioni, dei nessi insiti in frammenti di testo, percorrendo una mole straordinaria di informazioni, *tuttavia il dato rilevante è il ruolo che questa produzione viene ad assumere nelle diverse attività umane*. I LLM in quanto operano e interagiscono a livello di linguaggio naturale, avvalendosi di un campo di informazioni spaventosamente esteso, entrano con molta facilità in una molteplicità di attività umane basate sulla ricerca di informazioni,

elaborate in termini linguistici, e sulla produzione di messaggi in ambiti specifici. Si sperimentano quindi forme sempre più sofisticate di interrogazione, si cerca di circoscrivere la produzione di 'allucinazioni' ossia di falsi enunciati; del resto negli anni passati siamo vissuti entro la rivoluzione dei motori di ricerca, l'attività di googlare è diventata essenziale per qualsiasi operatività dal semplice plagio alla ricerca più sofisticata. Il problema della proprietà dei testi, delle conoscenze del loro utilizzo costituisce quindi un nodo ineludibile, portando ad un nuovo livello un problema posto e non risolto dai motori di ricerca e dai social networks sulla proprietà dell'informazione.

Nel frattempo le reti neurali, classificate rispetto vari tipi di 'addestramento' e auto-apprendimento possibili, stanno producendo risultati significativi in tutti i campi, vale per tutti l'esempio della scoperta di nuove varietà di molecole, in particolare proteine¹³.

Stiamo quindi parlando di un ecosistema tecnologico, connesso ad una costellazione di altri ecosistemi tecnologici che interviene sulla composizione sociale e sulla stratificazione delle diseguaglianze, rivoluzionando sistemi economici - dell'industria, della logistica, dei servizi intervenendo nella produzione culturale e nelle attività formative e via elencando, in generale nella produzione, condivisione e trasmissione delle conoscenze e delle *informazioni finalizzate ad uno scopo*.

Il senso generale dell'*Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence* dell'amministrazione USA è quindi quello di orientare ciò che si annuncia come un processo di straordinaria trasformazione in tutti i settori, di ridefinizione degli equilibri e di processi fondamentali della riproduzione sociale e del governo delle società.

Una serie di commenti all'E.O. si possono trovare nell'articolo 'Will the White House AI Executive Order deliver on its promises?'¹⁴, da parte di esperti della *Brookings Institution*.

I commenti che meritano una lettura attenta, sottolineano di volta in volta l'importanza di alcuni degli otto principi guida, approfondendone l'analisi, evidenziano la necessità di dare le necessarie competenze e risorse a tutti gli attori¹⁵ a cui si richiede di elaborare standard e di controllarne l'applicazione verificando l'efficacia delle procedure messa in atto. Rilevante è l'osservazione che evidenzia la mancanza di un approfondimento del ruolo delle istituzioni incaricate di regolare i mercati finanziari.

Se è vero che l'E.O. propone un modello di collaborazione tra tutti gli attori sulla scena dal livello governativo, alle grandi aziende che controllano di fatto lo sviluppo della tecnologia, alla società civile, ai cittadini, passando alle università e al mondo delle start up, collaborazione che si può idealmente fondare sul libero accesso alle conoscenze, *in un mondo fondato sulla più feroce competizione e appropriazione privata della conoscenza da parte degli hub dello sviluppo tecnologico* è poco probabile una dinamica realmente partecipativa.

Si evidenzia comunque come l'efficacia di qualsiasi misura si basi su una capillare ed estesa formazione che abiliti i cittadini e le comunità a comprendere la natura della innovazione tecnologica e le sue conseguenze¹⁶

Questa serie di commenti ed osservazioni si rivolgono specificamente al *Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence* in riferimento alla situazione degli Stati Uniti, ma pongono ovviamente questioni di ordine generale e costituiscono un utile contributo per l'approfondimento delle riflessioni che stiamo producendo sulle pagine della nostra rivista in merito all'ecosistema tecnologico dell'Intelligenza Artificiale, in relazione alla trasformazione complessiva

della formazione sociale globale, alle crisi ed i conflitti che la attraversano.

Dall'I.A. quale cambiamento antropologico?

“Ogni invenzione o tecnologia è un'estensione o un'autoamputazione del nostro corpo, che impone nuovi rapporti o nuovi equilibri tra gli altri organi e le altre estensioni del corpo. (...) In quanto estensione e accelerazione della vita sensoriale, ogni medium influenza contemporaneamente l'intero campo dei sensi, come spiegava tanto tempo fa il Salmista nel Salmo 115” Così scriveva Marshall McLuhan¹⁷ citato da Antonio Caronia(Caronia, Antonio. Virtuale. Mimesis Edizioni) il quale poi prosegue.

“Ma le tecnologie delle civiltà agricole e di quella industriale sono estensioni di singoli sensi: in particolare la tecnologia della stampa, che McLuhan considera fondamentale per la nascita del mondo moderno, è un'estensione della vista, e segna un predominio così forte di questo senso sugli altri da sconvolgere l'equilibrio del mondo precedente. Le tecnologie elettriche ed elettroniche estendono invece tutto il nostro sistema nervoso (“Con l'avvento della tecnologia elettrica l'uomo estese, creò cioè al di fuori di sé stesso, un modello vivente del sistema nervoso centrale.” Citazione da McLuhan) Il processo di esteriorizzazione dell'informazione accompagna lo sviluppo della civiltà.” Esteriorizzazione/interiorizzazione un doppio movimento insito nelle tecnologie di trasmissione e condivisione, trattazione dell'informazione.

.Il sistema dei social network con i suoi miliardi di utenti, ha già modificato culture, consuetudini e comportamenti, sentimenti veicolato emozioni, in generale è intervenuto sul complesso delle attività cognitive e relazionali. Lo straordinario flusso di informazioni che esso produce in ogni frazione di secondo, può essere interpretato messo a valore solo con l'uso delle diverse tecnologie di A.I.

Da anni i dispositivi di I.A. si stanno moltiplicando nel tessuto sociale e produttivo operando nel sistema delle reti digitali che connettono un sistema di nodi sempre più intelligenti e diffusi, sovrintendendo al loro sviluppo e configurazione, integrando verticalità e orizzontalità del loro governo.

Il summit di Londra e l'Executive Order di Biden indicano il fatto che, in questa diffusione e nell'incremento e differenziazione delle prestazioni dei dispositivi tecnologici, si sta superando una soglia oltre la quale i sistemi di regolazione sociale che hanno operato sino ad ora rischiano di non reggere, di non essere in grado di adempiere al proprio compito.

Ciò che viene esteriorizzato viene re-interiorizzato.

I cambiamenti intervengono ed interverranno sempre di più a livello micro e macro, creando estensioni, organi esterni a livello individuale, collettivo e sociale, contemporanei a modifiche profonde a livello psicologico, antropologico e culturale.

Indubbiamente una trasformazione così radicale produce sue retoriche, una sua ideologia che va ben oltre l'attività di marketing che gioca sull'ineluttabilità dell'innovazione tecnologica assieme all'orizzonte catastrofico.

Una ideologia che mette al centro e deve giustificare il processo di artificializzazione di componenti sempre più estese e di relazioni entro la rete di riproduzione delle forme di vita, degli ecosistemi e della società. Il processo di artificializzazione innesta l'innovazione tecnologica entro la crisi climatica e ambientale, entro la rottura degli ecosistemi, nella ingovernabilità delle formazioni sociali. Una ideologia, una visione del mondo che -nella promessa di un mondo nuovo reso possibile dall'innovazione tecnologica- nasconde le ragioni profonde delle crisi. Di contro un giusto

atteggiamento critico, una cultura del sospetto verso questa promessa di realizzazione del migliore dei mondi possibili, si trasforma in un rifiuto a prescindere -la cosiddetta logica del complotto- come abbiamo vissuto nel movimento no-vax durante la pandemia, che aveva come conseguenza l'incapacità di una critica radicale nei confronti di bigpharma, del controllo oligopolistico su tutta l'industria della salute.

D'altra parte, oltre al confronto sui social e qualche episodica manifestazione, resta inevasa la richiesta il bisogno di contare, partecipare, influenzare il corso degli eventi, di non essere pedine di un gioco che pochi altri da noi dirigono.

Queste considerazioni ci portano al nesso tra tecnologia dell'I.A. in particolare e democrazia¹⁸ ovvero la crisi sostanziale dei processi democratici, l'inadeguatezza della democrazia rappresentativa per cui ci si deve interrogare su quali forme di conflitto siano in grado di produrre nuove forme di condivisione delle conoscenze ed apprendimento collettivo, di auto-organizzazione, in grado di riprodursi impattando sui processi produttivi e le forme di governo; conflitti in grado di attraversare i confini delle stratificazioni sociali, culturali, territoriali e generazionali, in grado di affrontare le dinamiche della competizione geostrategica che riproducono nei vari contendenti logiche di dominio. I movimenti che nascono per contrastare alla radice le cause del riscaldamento globale, cambiamento climatico, della devastazione degli ecosistemi sono indubbiamente uno dei luoghi dove avviene questa riflessione critica, l'attraversamento dei confini, ma siamo solo all'inizio.

Se l'accento della riflessione è posto sull'attributo artificiale e sulla costellazione di termini e concetti collegati, l'approfondimento delle conoscenze sul mondo della vita e sulle dinamiche della capacità di produrre linguaggio e conoscenza è in realtà uno dei motori della grande trasformazione in corso e contribuisce peraltro alla costruzione di una visione, anche di una ideologia, di cosa caratterizzi in modo sostanziale le abilità linguistiche, cognitive e relazionali-conoscenza, emozioni e comportamento- dell'essere umano. La ricerca sull'origine di queste capacità e sulle differenze dell'uomo dalle altre specie, da i primati in particolare, rimette in gioco le visioni del mondo delle gerarchie e delle relazioni tra le diverse specie¹⁹.

Il cambiamento climatico e la devastazione degli ecosistemi sta generando quella che viene la sesta estinzione di massa, questa volta generata dall'azione antropica. Siamo dentro una transizione nella quale le vette della riproduzione artificiale dei meccanismi della conoscenza e della riproduzione della vita in realtà si toccano entro un mondo, un modello di riproduzione sociale, che si richiude su stesso come entro l'anello rotante di una grande astronave che i romanzi di fantascienza hanno prefigurato. *La grande astronave, il cyborg globale sembrano essere le metafore più adatte a descrivere il mondo che si sta prefigurando, ci manca la capacità di analisi, di prefigurazione e di narrazione di Antonio Caronia*. Salvo che nulla e nessuno garantisce che questa astronave sia capace di decollare ed una volta decollata non imploda per le sue contraddizioni e conflitti interni o che il cyborg -come nei migliori racconti- non impazzisca, elabori logiche dis-umane ovvero si frantumi in mille componenti in conflitto tra di loro.

In ogni caso sull'astronave non ci salgono tutti e tremende sono le differenze tra le classi di imbarco, quanto al cyborg non tutti possono connettersi e forti sono le differenze tra le diverse categorie di connessione.

Roberto Rosso

1. <https://www.politico.eu/article/sunak-the-influencer-how-the-uks-ai-summit-surprised-the-skeptics/> [↔]

2. <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023?fbclid=IwAR0oaVzzRtfGZSRAsKpIqvgKfzWxEltob4Q2SET11zvbWncRgJ7qElxcFU> [↔]
3. <https://hai.stanford.edu/news/reflections-foundation-models> <https://crfm.stanford.edu/> [↔]
4. Il Center for Research on Foundation Models (CRFM) dello Stanford Institute for Human-Centered Artificial Intelligence (HAI) ha coniato il termine “foundation model” nell’agosto 2021, riferendosi provvisoriamente a “qualsiasi modello addestrato su una ampia mole di dati (generalmente utilizzando l’aut-supervisione su larga scala) che può essere adattato (ad esempio, messo a punto) a un’ampia gamma di attività a valle”. Ciò si basava sulla loro osservazione che i termini, concetti parzialmente sovrapponibili esistenti non erano adeguati, sostenendo che “il ‘(grande) modello linguistico’ era troppo ristretto dato che l’attenzione non è solo sul linguaggio; ‘ modello auto-supervisionato” era troppo specifico per l’obiettivo di formazione; E il “modello pre-addestrato” ha suggerito che l’azione degna di nota è avvenuta dopo il “pre-addestramento”. Dopo aver considerato molti termini, hanno optato per il “ foundation model ” per enfatizzare la funzione prevista (cioè l’adattabilità a un successivo ulteriore sviluppo) piuttosto che la modalità, l’architettura o l’implementazione.[↔]
5. <https://arxiv.org/abs/2108.07258> [↔]
6. All actors have a role to play in ensuring the safety of AI: nations, international fora and other initiatives, companies, civil society and academia will need to work together. Noting the importance of inclusive AI and bridging the digital divide, we reaffirm that international collaboration should endeavour to engage and involve a broad range of partners as appropriate, and welcome development-orientated approaches and policies that could help developing countries strengthen AI capacity building and leverage the enabling role of AI to support sustainable growth and address the development gap.[↔]
7. In furtherance of this agenda, we resolve to support an internationally inclusive network of scientific research on frontier AI safety that encompasses and complements existing and new multilateral, plurilateral and bilateral collaboration, including through existing international fora and other relevant initiatives, to facilitate the provision of the best science available for policy making and the public good [↔]
8. *In doing so, we recognise that countries should consider the importance of a pro-innovation and proportionate governance and regulatory approach that maximises the benefits and takes into account the risks associated with AI. This could include making, where appropriate, classifications and categorisations of risk based on national circumstances and applicable legal frameworks.*[↔]
9. <https://artificialintelligenceact.eu/developments/>
https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence?at_campaign=20226-Digital&at_medium=Google_Ads&at_platform=Search&at_creation=RSA&at_goal=TR_G&at_advertiser=Webcomm&at_audience=ai%20act&at_topic=Artificial_intelligence_Act&at_location=IT&gclid=CjwKCAiA3aeqBhBzEiwAxFiOBsurwDHcEjofVBH6-qoqZIxQtoE114nuFesMOnph7LMD68MHXZA6CBoCDtYQAvD_BwE <https://www.coe.int/en/web/artificial-intelligence> [↔]
10. <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>
<https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/?fbclid=IwAR10M0l0PAbuxnEBzkh2wxuZROyWljHa3rWkmPD9Dp1eeFXBM-9scAg9CY#:~:text=The%20Executive%20Order%20establishes%20new,around%20the%20world,%20and%20more.> [↔]
11. <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/>
<https://www.whitehouse.gov/briefing-room/statements-releases/2023/02/16/fact-sheet-president-biden-signs-executive-order-to-strengthen-racial-equity-and-support-for-underserved-communities-across-the-federal-government/> [↔]
 - Catalyze AI research across the United States through a pilot of the National AI Research Resource—a tool that will provide AI researchers and students access to key AI resources and data—and expanded grants for AI research in vital areas like healthcare and climate change.
 - Promote a fair, open, and competitive AI ecosystem by providing small developers and entrepreneurs access to technical assistance and resources, helping small businesses commercialize AI breakthroughs, and encouraging the Federal Trade Commission to exercise its authorities.[↔]
13. <https://www.wired.it/article/intelligenza-artificiale-molecole-proteine-nuovi-farmaci-studio/>
<https://www.agendadigitale.eu/sanita/alphafold-come-lintelligenza-artificiale-rivoluziona-la-biologia/> [↔]
14. <https://www.brookings.edu/articles/will-the-white-house-ai-executive-order-deliver-on-its-promises/> [↔]

15. While the EO specifically calls out these and other points in the design and deployment of AI, more clarity is needed. That is, how is a civil rights violation identified in opaque AI systems, and who decides if the situation presents itself as worthy of punitive action? Additionally, what will be the recourse for individuals harmed by discriminatory AI systems? On these points, Congress will definitely have to provide guidance to federal agencies. Who is seated at the table in the design and deployment of AI also matters, especially the inclusion of academic, or civil society experts that understand the lived experiences of communities (including their trauma) who have become the primary subjects of existing and emerging technologies.
(...)
Legislation will be needed to strengthen the capabilities of existing agencies outside the national security area to meet the challenges of AI. Other than a new privacy law that would protect kids, the EO does not identify where the administration needs new authority to regulate AI effectively. This is a missed opportunity.
[←]
16. As the United States moves closer to regulating AI, educating the general public about the limitations, benefits, and harms of AI will elevate the role of everyday Americans in effective oversight and empower them to become active stakeholders. [←]
17. *Marshall McLuhan, Gli strumenti del comunicare [Understanding Media, 1964], trad. it. di Ettore Capriolo, Garzanti, Milano 1986 (1^ ed. Il Saggiatore, 1967), pp. 64/65.*[←]
18. La seconda conferenza sull'intelligenza artificiale: A.I. e democrazia
<https://transform-italia.it/note-sulla-conferenza-sullintelligenza-artificiale/> [←]
19. Rispetto a queste due questioni fondamentali vale la pena di leggere Alain Ehrenberg *La meccanica delle passioni - Cervello, comportamento, società* -2019 Piccola Biblioteca Einaudi ed. originale del 2018 Raymond Corbey *Metafisiche delle scimmie* Negoziando il confine animali-umani a cura di Paola cavalieri 2008 Bollati Boringhieri ed. originale 2005[←]