



Piero Ugliengo<sup>a</sup>, Claudio Greco<sup>b</sup>, Marco De Vivo<sup>c</sup>

<sup>a</sup>Dipartimento di Chimica - Università di Torino

piero.ugliengo@unito.it

<sup>b</sup>Dipartimento di Scienze dell'Ambiente e della Terra - DISAT - Università degli Studi di Milano Bicocca

claudio.greco@unimib.it

<sup>c</sup>Laboratory of Molecular Modelling and Drug Discovery, Istituto Italiano di Tecnologia

marco.devivo@iit.it

# ARTIFICIAL INTELLIGENCE AND MODELING FOR CHEMISTRY

*Dal 26 al 30 agosto 2024 si è svolta a Milano la XXVIII edizione del Congresso Nazionale della SCI, all'interno del quale è stata organizzata la giornata tematica sull'intelligenza artificiale e modellistica per la chimica che ha coinvolto le Divisioni di Chimica Fisica, Chimica Teorica e Computazionale, Chimica Farmaceutica, Chimica per le Tecnologie, Chimica degli Alimenti, nonché il Dipartimento di Scienze Chimiche e Tecnologie dei Materiali del Consiglio Nazionale delle Ricerche.*

**N**ella settimana 26-30 agosto 2024 si è svolta a Milano la **XXVIII edizione del Congresso Nazionale della Società Chimica Italiana** con una partecipazione molto elevata. L'organizzazione del congresso prevedeva, oltre alle giornate dedicate ai lavori specifici delle varie divisioni della SCI, una **giornata focalizzata su temi di interesse trasversale**. Un tema ha riguardato il ruolo dell'intelligenza artificiale e della modellistica molecolare nelle scienze chimiche con il coinvolgimento delle Divisioni di Chimica Fisica, Chimica Teorica e Computazionale, Chimica Farmaceutica, Chimica per le Tecnologie e Chimica degli Alimenti, nonché il Dipartimento di Scienze Chimiche e Tecnologie dei Materiali del Consiglio Nazionale delle Ricerche, così da garantire la massima trasversalità tra diverse linee di ricerche accumulate dal ruolo sempre più importante di strumenti legati all'intelligenza artificiale (AI). La AI ha avuto, come dimostrano numerosissimi articoli sia sui quotidiani che sui social media generalisti, un enorme impatto in tutte le attività del mondo produttivo non solo industriale e nell'educazione scolastica di ogni grado. La recente apertura alla collettività di strumenti basati sui modelli di linguaggio di grandi dimensioni (LLM) come OpenAI-ChatGPT, Anthropic-Claude e Google DeepMind-Gemini stanno rivoluzionando ogni aspetto della società. Osservatori esperti predicono che il loro impatto e, specialmente, la

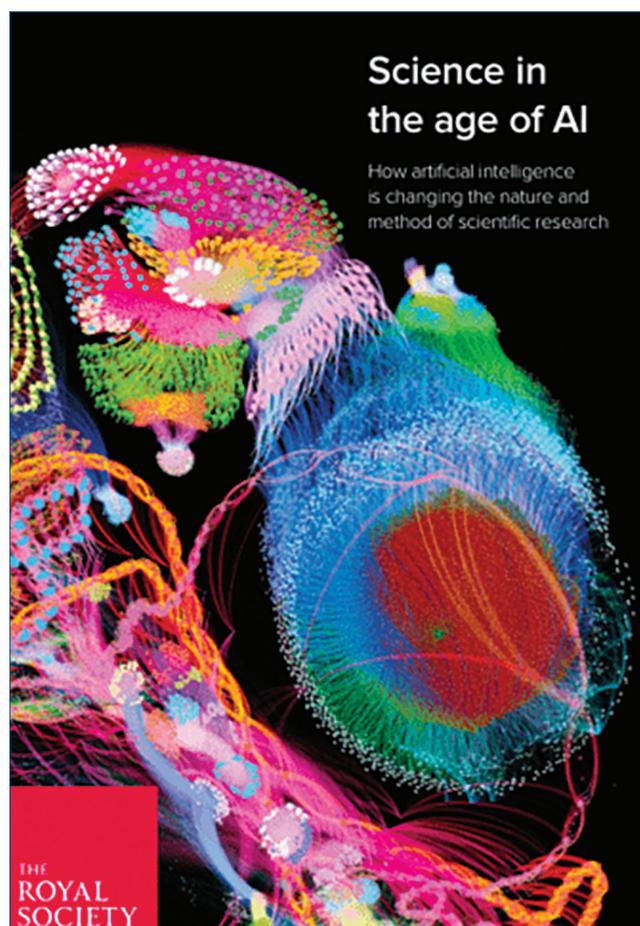


Fig. 1 - Frontespizio del documento della Royal Society sull'intelligenza artificiale nella scienza

velocità della loro adozione sarà decisamente superiore alle precedenti rivoluzioni industriali o della stessa rete internet. La scienza non è certo esente da questo processo. È significativo che la Royal Society abbia pubblicato un documento aperto e corposo dal titolo: **Science in the age of AI - how artificial intelligence is changing the nature and method of scientific research** (Fig. 1). Gli scienziati e i docenti a fronte di nuove rivoluzioni hanno il compito di stimolarne lo studio approfondito per capirne le profonde implicazioni e non essere schiavi di catastrofismi dettati dalla naturale tendenza di rifiutare le novità per proteggere la propria *comfort zone*.

Ovviamente anche la chimica è stata fortemente permeata dalla AI, specialmente in questo ultimo decennio. Bisogna però ricordare che l'introduzione di modelli matematici, già basati sulle reti neurali, in ambito chimico era già presente negli anni Novanta con il classico lavoro di J. Gastgeiger e J. Zupan [1]. D'altra parte, negli stessi anni, le reti neurali (NN) in chimica erano limitate dalla potenza computazionale e usavano architetture semplici come i multi-layer perceptron (MLP). L'allenamento avveniva con metodi del gradiente minimo e della *backpropagation* di base, su piccole basi di dati. Oggi, grazie alle elevatissime potenze dei Graphic Processing Units (GPU) e di framework informatici avanzati (TensorFlow, PyTorch) disponibili liberamente presso la comunità scientifica, vengono utilizzate reti profonde (DNN), convoluzionali (CNN) e basate sui grafi (GNN). I metodi di ottimizzazione

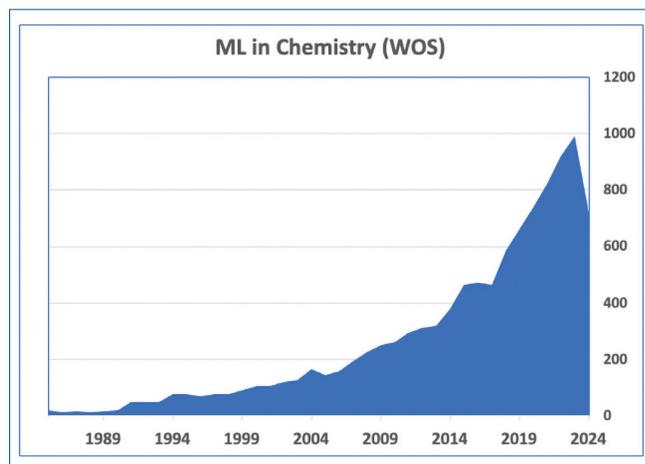


Fig. 2 - Risultato della ricerca per Topic sul Web of Science della stringa "ML in Chemistry" alla data del 19/08/2024

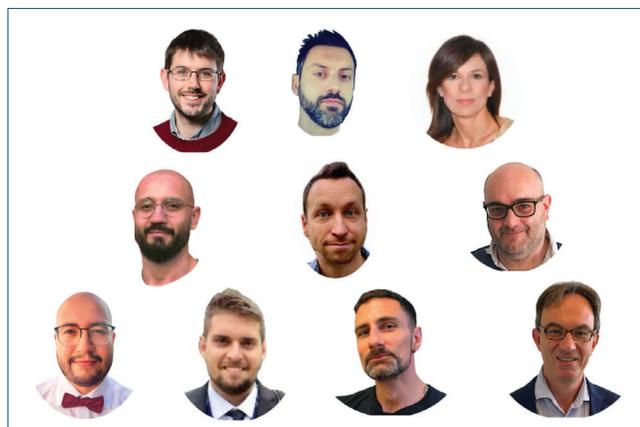


Fig. 3 - I keynote speaker: (in alto) Michele Ceriotti, Alfonso Pedone, Maria Cristina De Rosa; (al centro) Giovanni Maria Piccini, Massimo Delle Piane, Piero Macchi; (in basso) Manuel Jose Ruiz Munevar, Eugenio Alladio, Luca Dellafiara e Andrea Cavalli

sono più sofisticati (Adam, dropout) e l'accesso a grandi database chimici (PubChem, ChEMBL) permette di applicare le NN alla scoperta di farmaci, materiali, e simulazioni chimiche con precisione sempre più elevata. Se interroghiamo ChatGPT relativamente al ruolo del machine learning (ML) in chimica otteniamo: *"Il machine learning (ML) nelle scienze chimiche è l'applicazione di algoritmi di apprendimento automatico per analizzare, modellare e prevedere comportamenti chimici complessi. Viene utilizzato per accelerare la scoperta di nuovi materiali, ottimizzare reazioni chimiche, prevedere proprietà molecolari e analizzare grandi quantità di dati sperimentali e computazionali. Il ML permette di identificare pattern e correlazioni che possono non essere evidenti tramite metodi tradizionali, riducendo il tempo e i costi della ricerca chimica e migliorando l'efficienza dei processi"*. La rapidissima crescita di queste metodiche è evidente dalla ricerca del topic "ML in Chemistry" sul Web of Science come illustrato nella Fig. 2.

La giornata tematica è stata organizzata in ben nove *keynotes* su invito di 30' a cui hanno contribuito colleghi italiani dall'elevato profilo scientifico internazionale e otto contributi orali di 15' selezionati tra giovani ricercatori emergenti nel campo del ML e della modellista teorica in chimica (Fig. 3). Nel seguito verranno sintetizzati i soli interventi delle *keynote* per ragioni di spazio. L'apertura della **sessione mattutina** (Chair: Piero Ugliengo) è stata affidata a Michele Ceriotti, che dirige il labora-



torio di Scienza e Modellazione Computazionale, presso l'Istituto dei Materiali dell'EPFL, relativo allo sviluppo di metodi per la modellazione atomistica dei materiali basati sulla meccanica statistica e sull'apprendimento automatico. Ceriotti è internazionalmente noto per i suoi studi fondamentali del ML in chimica. Nel suo contributo "*Bottom-up machine learning for chemical modeling*" sono stati illustrati i più moderni metodi di apprendimento da dati quanto-meccanici rigorosi. Le applicazioni ora possibili con questi strumenti permettono un'estensione importante degli intervalli di temperatura, pressione e tempo di molti sistemi di grande interesse applicativo [2]. Alfonso Pedone, professore associato del Dip. di Scienze Chimiche e Geologiche dell'Università di Modena e Reggio-Emilia nel contributo "*Enhancing atomistic simulations of oxide glasses through the power of machine learning*" ha mostrato l'uso del ML in materiali vetrosi ossidici a composizione chimica variabile, che ora consentono lo studio rigoroso delle proprietà spettroscopiche (NMR) di questi importanti materiali [3]. Maria Cristina De Rosa dirigente di ricerca della sezione di Roma, presso l'Università Cattolica, dell'Istituto di Scienze e Tecnologie Chimiche "Giulio Natta" (SCITEC) del CNR nel contributo "*From molecular modeling to data science: shaping computer-aided drug design for new therapeutic strategies*" ha mostrato il ruolo del ML nella modellazione molecolare e progettazione di farmaci [4]. La **seconda sessione mattutina** (Chair: Bartolomeo Civalleri) ha visto il contributo di Giovanni Maria Piccini "*Unravelling the molecular aspects of catalysts under operating conditions combining simulations and machine learning*"; il prof. Piccini, già all'Università di Groningen e ora professore associato del Dip. di Scienze Chimiche e Geologiche dell'Università di Modena e Reggio-Emilia, nel suo contributo ha mostrato come l'apprendimento automatico da dati *ab initio* di dinamica molecolare possa consentire l'esplorazione molto più accurata della superficie di potenziale così cruciale nei processi di catalisi eterogenea che avvengono nelle zeoliti [5]. Massimo Delle Piane, ricercatore presso il Dipartimento di Scienze e Tecnologie Applicate del Politecnico di Torino nel contributo "*The hidden dynamics of metals: a deep dive fueled by ML*" ha mostrato come il ML grazie ad evoluzioni temporali molto lunghe riveli

una natura dinamica delle superfici e dei clusters metallici ad alta temperatura destinato a cambiare il paradigma di staticità attribuita ai sistemi metallici nelle applicazioni della modellistica tradizionale [6]. Piero Macchi, professore ordinario presso il Dipartimento di Chimica, Materiali e Ingegneria Chimica del Politecnico di Milano con una lunga esperienza di cristallografo, nel contributo "*Quantum crystallography and quantum computing*" ha illustrato i concetti fondamentali della cristallografia quantistica che combina informazioni sperimentali provenienti da misure di diffrazione a raggi X sui cristalli e la teoria quantistica mostrandone le potenziali applicazioni nel quantum computing [7]. Nella **prima parte della sessione pomeridiana** (Chair: Marco De Vivo) Manuel Jose Ruiz Munevar, ricercatore presso il Molecular Modeling & Drug Discovery Laboratory dell'Istituto Italiano di Tecnologia, nel contributo "*Deep learning to design novel and selective nkcc1 inhibitor chemotypes for the treatment of brain disorders with defective nkcc1/kcc2 ratio*" ha illustrato come diversi studi mostrino un'alterata omeostasi del cloruro nei disturbi cerebrali. L'inibizione della proteina di membrana NKCC1 migliora i sintomi, portando alla scoperta dell'inibitore IAMA-6 e di nuovi composti con specifiche proprietà sviluppati utilizzando metodi computazionali e di ML [8].

Eugenio Alladio, professore associato del Dipartimento di Chimica dell'Università di Torino, nel contributo "*Chemometrics in clinical chemistry: examples of open-source applications for method validation and evaluation of real cases*" ha mostrato come in chimica clinica, la fusione della chemiometria con il software open-source SpectrApp (di cui è autore) sia fondamentale per garantire risultati analitici solidi e replicabili, in particolare nella validazione dei metodi e nelle valutazioni di casi reali [9]. Luca Dellafiora, ricercatore presso il Dipartimento di Scienze degli Alimenti e del Farmaco dell'Università di Parma nel contributo "*Food toxicology and food bioactives assessment through in silico methods - is AI the new era of testing?*" ha mostrato come AI e il ML possono migliorare la sicurezza alimentare consentendo analisi multidimensionali su larga scala, affinando la valutazione del rischio chimico e facilitando il processo decisionale attraverso algoritmi avanzati [10]. La parte

### Commento sui premi Nobel 2024

A poco più di un mese dal termine del congresso della SCI, l'Accademia delle Scienze Svedese ha assegnato entrambe i premi Nobel per la Fisica e la Chimica a temi strettamente connessi a quelli della giornata tematica: "Artificial intelligence and modeling for chemistry" i cui dettagli sono stati descritti in questo articolo. Il premio Nobel per la Fisica è stato assegnato a John J. Hopfield dell'Università di Princeton e a Geoffrey E. Hinton dell'Università di Toronto "Per le scoperte e invenzioni fondamentali che permettono l'apprendimento automatico con reti neurali artificiali". I vincitori hanno utilizzato strumenti della fisica per costruire metodi che hanno contribuito a porre le basi per l'apprendimento automatico odierno (ML: Machine Learning). John Hopfield ha creato una struttura capace di memorizzare e ricostruire informazioni, mentre Geoffrey Hinton ha inventato un metodo che scopre autonomamente le proprietà nei dati, diventato cruciale per le grandi reti neurali artificiali in uso. Il premio Nobel per la Chimica (v. commento a pag. 10 di questo numero) è stato assegnato a David Baker dell'Università di Washington "per la progettazione computazionale delle proteine" e Demis Hassabis e John M. Jumper di Google DeepMind "per la previsione della struttura proteica" utilizzando l'intelligenza artificiale. David Baker è riuscito nell'impresa quasi impossibile di costruire tipi di proteine completamente nuovi specificandone prima la struttura tridimensionale desiderata per poi predire, *in silico*, quali amminoacidi sarebbero stati necessari per la realizzazione della struttura desiderata, grazie alle informazioni sperimentali contenute nel *protein data bank*. Demis Hassabis e John Jumper hanno sviluppato un modello di intelligenza artificiale per risolvere un problema vecchio di 50 anni: prevedere il folding delle proteine, cruciale per l'espressione del loro funzionamento. Così mentre con il software ROSETTA, messo a punto da David Baker, è possibile predire la sequenza primaria estratta dalla ipotetica struttura proteica e farla poi codificare da un batterio geneticamente modificato, Demis Hassabis e John M. Jumper con la creazione modello AI AlphaFold2 sono riusciti a prevedere la struttura di quasi tutte le 200 milioni di proteine identificate dai ricercatori, partendo esclusivamente dalla loro sequenza primaria. Dalla loro scoperta, AlphaFold2 è stato utilizzato da oltre due milioni di persone di 190 Paesi. Nella giornata tematica si possono identificare specifici contributi basati sul ML in vari settori della chimica. Si parte dall'illustrare le sofisticate implementazioni delle più moderne reti neurali profonde in ambito chimico (M. Ceriotti), alla predizione di spettri NMR in materiali vetrosi (A. Pedone), alla catalisi eterogena (G.M. Piccini) fino ad applicazioni in campo del drug-design (M.C. De Rosa) e dell'inibizione di canali di membrana (M.J. Ruiz Munevar). È perciò facile immaginare che la comunità chimica tutta sarà stimolata a esplorare l'utilizzo di strumenti di AI nell'immediato futuro che, ben lontani dal rimpiazzare la creatività umana, al contrario ne potenzieranno l'effetto grazie a scoperte più rapide in campi cruciali per la nostra sopravvivenza, dai materiali per l'energia allo sviluppo di rivoluzionari farmaci.



Fig. 4 - La Prof.ssa Maria Cristina Menziani, nel momento della consegna della Medaglia Cesare Pisani della Società Chimica Italiana da parte del Presidente Prof. Gianluca Farinola

**conclusiva pomeridiana** (Chair: Claudio Greco) ha visto la sessione dedicata alla *Medaglia Cesare Pisani*, conferita dalla Società Chimica Italiana alla Prof.ssa Maria Cristina Menziani, per “l’eccellente attività scientifica volta alla razionalizzazione ed interpretazione di dati sperimentali riguardanti sistemi di interesse biologico e farmacologico, basate sullo sviluppo di innovative strategie di modellistica molecolare funzionali alla progettazione di nuovi farmaci e biomateriali, contribuendo significativamente al progresso della ricerca in campo terapeutico”. La Prof.ssa Menziani, nel suo contributo: “1988-2024, a 35-years journey across computational life and material sciences” ha illustrato il suo trentennale percorso di ricerca mettendo in luce come la chimica computazionale e la modellistica possano contribuire tramite gli stessi strumenti a chiarire sia le delicate interazioni tra un farmaco e il suo recettore che le caratteristiche superficiali di complessi materiali a base inorganica utilizzabili come biomateriali, unendo quindi la *soft-matter* con la *hard-matter* [11] (Fig. 4).

Andrea Cavalli, professore ordinario presso il Dipartimento di Farmacia e Biotecnologie dell’Università di Bologna e associato all’IIT e attuale direttore del CEAM nel contributo “*Role of computational methods in the era of precision medicine*” ha mostrato come AI e ML possano accelerare la dinamica molecolare nella scoperta dei farmaci, consentendo un’esplorazione efficiente degli spazi conformazionali delle molecole, migliorando i cal-

coli dell’energia libera e della cinetica e affrontando le sfide dello sviluppo di nuovi farmaci [12]. In conclusione, la sessione sull’intelligenza artificiale e modellistica per la chimica ha registrato un’ampia partecipazione di colleghi con interessanti contributi da parte di giovani ricercatrici e ricercatori. Anche la fase di discussione a valle di ogni contributo è stata rilevante.

## BIBLIOGRAFIA

- [1] J. Gasteiger, J. Zupan, *Angew. Chem. Int. Ed. Engl.*, 1993, **32**, 503.
- [2] S. Back, A. Aspuru-Guzik *et al.*, *Digital Discovery*, 2024, **3**, 23.
- [3] M. Bertani, T. Charpentier *et al.*, *J. Chem. Theory Comput.*, 2024, **20**, 1358.
- [4] C. Camponeschi, B. Righino *et al.*, *Biomolecules*, 2023, **13**, 1047.
- [5] G.M. Piccini, M.S. Lee *et al.*, *Catal. Sci. Technol.*, 2022, **12**, 12.
- [6] M. Cioni, M. Delle Piane *et al.*, *Adv. Science*, 2024, **11**, 2307261.
- [7] P. Macchi, *Cryst. Rev.*, 2020, **26**, 209.
- [8] M.J. Ruiz Munevar, V. Rizzi *et al.*, *J. Am. Chem. Soc.*, 2024, **146**, 552.
- [9] E. Alladio, F. Trapani *et al.*, *J. Pharm. and Biom. Anal.* 2024, **244**, 116113.
- [10] L. Dellafiora, G. Aichinger *et al.*, *Food chemistry*, 2019, **270**, 61.
- [11] D. Carrozza, G. Malavasi *et al.*, *Int. J. Mol. Sci.*, 2023, **24**, 880.
- [12] S. Majumdar, F. Di Palma *et al.*, *J. Chem. Inform. Model.*, 2023, **63**, 4814.

## XXVIII National Congress of the Italian Chemical Society

From Aug. 26-30, 2024, the XXVIII National Congress of the Italian Chemical Society (SCI) was held in Milan, Italy, within which the thematic day on artificial intelligence and modeling for chemistry was organized involving the SCI divisions of Physical Chemistry, Theoretical and Computational Chemistry, Pharmaceutical Chemistry, Chemistry for Technologies, Food Chemistry, and the National Research Council - CNR.