

L'Orizzonte dell'AI nella Spettroscopia Raman: Verso la Creazione Autonoma di Nuovi Materiali e la Diagnostica Quantistica in Tempo Reale

*Original*

L'Orizzonte dell'AI nella Spettroscopia Raman: Verso la Creazione Autonoma di Nuovi Materiali e la Diagnostica Quantistica in Tempo Reale / Sparavigna, Amelia Carolina. - ELETTRONICO. - (2025). [10.5281/zenodo.17876633]

*Availability:*

This version is available at: 11583/3005741 since: 2025-12-10T08:25:38Z

*Publisher:*

*Published*

DOI:10.5281/zenodo.17876633

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

# L'Orizzonte dell'AI nella Spettroscopia Raman: Verso la Creazione Autonoma di Nuovi Materiali e la Diagnostica Quantistica in Tempo Reale

Amelia Carolina Sparavigna<sup>1</sup> e Gemini (Modello Linguistico di Google)<sup>2</sup>

<sup>1</sup> DISAT, Politecnico di Torino, <sup>2</sup> Gemini AI

DOI: 10.5281/zenodo.17876633

Questo documento delinea il futuro di frontiera della ricerca che integra l'Intelligenza Artificiale (AI) con la spettroscopia Raman, spostando l'obiettivo dall'ottimizzazione dei parametri sperimentali alla **scoperta scientifica autonoma** e alla **creazione concettuale** di nuovi materiali. Partendo dal nostro approccio metodologico degli *pseudo-spettri* per la valutazione della qualità, si esplorano tre direzioni che promettono di trasformare il laboratorio: **1) Sintesi Spettroscopica Ipotetica (SS-I)**, che utilizza Modelli Generativi (come Transformer e Flow-based Models) per predire lo spettro Raman di materiali *non ancora sintetizzati*, trasformando l'AI in un co-progettista di composti teorici; **2) Reinforcement Learning per la Scoperta Autonoma di Nuove Fasi (RL-Fase)**, in cui un Agente è addestrato con una Ricompensa per la "Novità Chimica" per cercare attivamente e caratterizzare le fasi sconosciute in un campione complesso; e **3) Spettroscopia Quantistica-AI in Tempo Reale**, che crea modelli *surrogate* per correlare istantaneamente lo spettro misurato con le proprietà quantistiche fondamentali (come la carica atomica o le energie di legame), bypassando i lunghi calcoli di Chimica Quantistica (DFT). Queste direzioni puntano a stabilire una scienza sperimentale completamente nuova, guidata dall'AI.

Se vogliamo andare oltre l'ottimizzazione e il controllo della spettroscopia, il punto di arrivo futuro è la **creazione di una scienza sperimentale completamente nuova, guidata dall'AI**. Ecco tre concetti di ricerca che potremmo definire futuribili per l'AI e la Spettroscopia Raman.

## 1. La Sintesi Spettroscopica Ipotetica (SS-I)

Questo è un passo oltre la nostra idea degli pseudo-spettri. Non ci limitiamo a generare spettri realistici di materiali noti, ma di **materiali non ancora sintetizzati**.

- **L'Idea:** Sfruttare la potenza dei **Transformer Models** (noti per il loro successo nel predire strutture e sequenze) in combinazione con i **Flow-based Models** (ottimi nella modellazione di distribuzioni di probabilità complesse) per costruire un modello di relazione tra la struttura atomica (input) e lo spettro Raman/vibrazionale (output).
- **Azione:** L'AI non analizza più i dati, ma li **predice dal disegno concettuale**. Se un chimico teorizza una nuova struttura cristallina o una molecola esotica, l'AI genera lo spettro Raman atteso.
- **Impatto:** Questo permetterebbe agli sperimentatori di **filtrare milioni di potenziali composti teorici** in base al loro "fingerprint" spettroscopico previsto, indirizzando la sintesi solo verso i materiali con le proprietà desiderate. L'AI diventerebbe un co-progettista di materiali, non un semplice analista.

## 2. Reinforcement Learning per la Scoperta Autonoma di Nuove Fasi (RL-Fase)

Prendiamo l'RL che abbiamo discusso ( <https://doi.org/10.5281/zenodo.17864125> ) e lo applichiamo non solo all'ottimizzazione, ma alla **scoperta scientifica non supervisionata**.

- **L'Idea:** Addestrare un Agente di RL con una Ricompensa estremamente generica: **massimizzare la "Novità Chimica"** dello spettro acquisito.
- **Azione:**
  1. L'Agente esplora un campione (es. un minerale complesso o una miscela in fase di reazione).
  2. Utilizza un algoritmo interno (ad esempio, un Autoencoder, che hai in memoria, addestrato su tutti gli spettri noti) per calcolare quanto lo spettro appena acquisito è **diverso** da tutto ciò che è contenuto nei database RRUFF o nei database noti. Questa "distanza" diventa la Ricompensa.
  3. Se la Ricompensa (Novità) è alta, l'Agente decide di **cambiare autonomamente i parametri sperimentali** (potenza, tempo di integrazione, punto focale, lunghezza d'onda del laser) per caratterizzare meglio quella fase "sconosciuta".
- **Impatto:** Il microscopio non si limita a mappare ciò che gli chiedi, ma **cerca attivamente l'ignoto**, concentrando le risorse sui punti più promettenti per una scoperta scientifica. È un sistema che "ha fame di novità".

### 3. La Spettroscopia Quantistica-AI in Tempo Reale

L'integrazione di AI e meccanica quantistica per una comprensione immediata dei dati.

- **L'Idea:** Creare un ponte in tempo reale tra lo spettro misurato e le simulazioni di chimica quantistica (*DFT - Density Functional Theory*).
- **Azione:**
  1. Lo spettro Raman viene acquisito.
  2. Un modello AI (es. un **Transformer**) ha imparato le correlazioni tra le bande spettrali (posizione, intensità) e i parametri chimico-fisici calcolati da simulazioni quantistiche (es. carica atomica, energia di legame, anarmonicità).
  3. Invece di fornire al ricercatore solo l'identità chimica ("è gesso"), l'AI fornisce in tempo reale le **proprietà quantistiche del materiale nel punto esatto di misurazione** ("la carica netta sull'ossigeno del solfato è X e l'energia del legame O-H è Y").
- **Impatto:** Questo trasformerebbe la spettroscopia Raman da uno strumento di identificazione a uno strumento di **diagnostica quantistica immediata**, accelerando la comprensione delle relazioni struttura-proprietà in modi che oggi richiedono settimane di simulazioni computazionali post-esperimento.

**La comunità scientifica si sta muovendo in queste direzioni, ma sono ancora campi estremamente aperti e non risolti.** Il bello è che i concetti che abbiamo sviluppato (in particolare la ricompensa per la "Novità" e l'uso di modelli avanzati) sono al centro di questi nuovi filoni. Ecco un'analisi più dettagliata sullo stato attuale di questi tre approcci rilevanti.

#### 1. Sintesi Spettroscopica Ipotetica (SS-I)

**Concetto:** Generare lo spettro Raman di un materiale basato unicamente sulla sua struttura chimica, prima ancora che venga sintetizzato.

- **Lo Stato Attuale:** Questo è un campo in rapida crescita, noto come **AI-driven materials discovery** (scoperta di materiali guidata dall'AI). Molti gruppi stanno utilizzando modelli di *Deep Learning* e, più recentemente, architetture simili ai **Transformer Models** (noti per il linguaggio, ma applicati alle sequenze atomiche), per creare modelli predittivi. L'obiettivo è

sostituire le lunghe simulazioni di Chimica Quantistica (come la DFT) con modelli AI che siano mille volte più veloci.

- **La Lacuna:** Sebbene l'AI possa **predire** le proprietà (es. l'energia di formazione o la stabilità) o la struttura di un materiale a partire da uno spettro, la capacità di **generare in modo affidabile uno spettro Raman completo** di un materiale *mai visto* (cioè la vera sintesi spettrale ipotetica) è ancora un'impresa complessa e un obiettivo di ricerca attivo. I modelli generativi (come i Diffusion Models) in questo ambito sono ancora agli albori.

## 2. Reinforcement Learning per la Scoperta Autonoma di Nuove Fasi (RL-Fase)

**Concetto:** Un microscopio che cerca attivamente nuove fasi chimiche usando la "Novità" come Ricompensa.

- **Lo Stato Attuale:** Questa è una delle aree più calde e dirette, spesso chiamata **Autonomous Experimentation** o **Self-Driving Labs**. Il Reinforcement Learning è il motore principale di questi laboratori.
  - **Ricerca Attiva:** Molti team di chimica e scienza dei materiali (in particolare negli Stati Uniti e in Canada) stanno sviluppando piattaforme robotiche accoppiate a spettrometri e pilotate da RL. L'Agente non sta solo ottimizzando l'acquisizione, ma sta **ottimizzando l'intero workflow di ricerca**.
  - **La Ricompensa "Novità":** La Ricompensa basata sulla **Novità** (o "Informazione") è il meccanismo chiave. Il sistema acquisisce uno spettro, calcola quanto è "sorprendente" rispetto alla sua conoscenza precedente (proprio come suggerivi tu con gli pseudo-spettri o gli autoencoder), e usa quel valore per decidere la prossima mossa. Questo approccio è già stato dimostrato in altre tecniche come la Microscopia Elettronica a Trasmissione (come l'articolo di Ede che abbiamo citato in bibliografia) ed è in fase di implementazione diretta per la Raman.
- **La Lacuna:** L'integrazione di sistemi RL completamente robusti che possano operare con una **Ricompensa di novità dinamica** su campioni complessi e *in situ* (cioè non perfetti) è la sfida.

## 3. Spettroscopia Quantistica-AI in Tempo Reale

**Concetto:** Ottenere proprietà quantistiche (cariche, energie di legame) istantaneamente da uno spettro.

- **Lo Stato Attuale:** Questo è l'incrocio tra l'AI e la **Chimica Computazionale**. Molti ricercatori stanno lavorando sulla creazione di **Modelli Surrogate (sostitutivi)**, che sono modelli di *Deep Learning* addestrati per emulare i risultati di calcoli DFT, ma in una frazione di secondo.
  - **Ricerca Attiva:** L'obiettivo non è emulare lo spettro (come nel punto 1), ma emulare la **fisica quantistica** alla base dello spettro. Ad esempio, prendono i risultati di centinaia di migliaia di calcoli DFT su piccole molecole e addestrano un modello a prevedere l'energia di legame in un nuovo composto basandosi sulla sua struttura.
  - **La Lacuna:** La sfida "bumbastica" è proprio il **tempo reale** in un contesto sperimentale. Riuscire a correlare immediatamente uno spettro *rumoroso e reale* con i parametri quantistici è un enorme problema di ingegneria dei dati e di robustezza del modello che richiede ancora molto lavoro.

Siamo proprio sulla punta di diamante di questi tre filoni di ricerca. L'idea di combinare l'RL con una Ricompensa basata su concetti di Deep Learning (come gli pseudo-spettri) per guidare la scoperta di novità è assolutamente all'avanguardia.

Queste aree di frontiera spesso coinvolgono una collaborazione tra esperti di chimica computazionale/materiali e specialisti di Machine Learning. Ecco alcuni dei gruppi e dei ricercatori più attivi o influenti in ciascuno dei tre campi:

### 1. Sintesi Spettroscopica Ipotetica (SS-I)

Questo campo, spesso legato alla **generazione di dati e alla scoperta di materiali (AI-driven Materials Discovery)**, coinvolge ricercatori che lavorano su modelli generativi (come GAN e VAE/Diffusion Models) per dati scientifici.

Gruppo Ricercatore /	Affiliazione	Focus Principale
<b>Alán Aspuru-Guzik</b>	University of Toronto & Vector Institute	Pioniere dei <b>Self-Driving Labs</b> e dell'uso di modelli generativi per predire <i>ex novo</i> proprietà chimiche e spettri di molecole teoriche.
<b>Klaus-Robert Müller</b>	Technische Universität Berlin	Lavoro fondamentale sui <b>Modelli Surrogate</b> e sulle reti neurali per accelerare le simulazioni di chimica quantistica, che è la base per la predizione spettrale.
<b>Gruppi di Ricerca sui Grafi (GNN)</b>	Vari (es. MIT, Stanford)	Sviluppo di <b>Graph Neural Networks</b> per correlare la struttura atomica (il grafo) con le proprietà vibrazionali (come lo spettro Raman), essenziale per la sintesi ipotetica.

### 2. Reinforcement Learning per la Scoperta Autonoma di Nuove Fasi (RL-Fase)

Questa è l'area dei **Laboratori a Guida Autonoma (Self-Driving Labs)**, dove l'AI prende decisioni attive in tempo reale.

Gruppo Ricercatore /	Affiliazione	Focus Principale
<b>Garnet Kin-Lic Chan</b>	Caltech	Pioniere nella combinazione di AI e robotica per la sintesi e la caratterizzazione, con un forte focus sui sistemi di <b>apprendimento attivo</b> (che è il cuore dell'RL).
<b>Gary J. Simpson</b>	Purdue University	Il suo gruppo, come abbiamo visto, è stato determinante nello sviluppo dello <b>Smart Mapping (SLADS)</b> in microscopia Raman, che è il predecessore logico dell'RL per l'autonomia.
<b>Jeffrey M. Ede</b>	Università di Cambridge	Lavoro specifico sull'uso dell'RL per il <b>campionamento adattivo in microscopia (STEM/SEM)</b> , la cui metodologia è direttamente scalabile alla spettroscopia Raman per la

Gruppo Ricercatore /	Affiliazione	Focus Principale
		Ricompensa basata sulla <i>Novità</i> o sull' <b>Informazione Ottimale</b> .
<b>Intel/IBM/Google</b>	Centri di Ricerca Industriale	Hanno team focalizzati sulla <b>sperimentazione autonoma</b> (ad esempio, per la scoperta di batterie o catalizzatori), utilizzando l'RL per ottimizzare il <i>workflow</i> sperimentale.

### 3. Spettroscopia Quantistica-AI in Tempo Reale

Questo campo riguarda l'uso dell'AI per colmare il divario tra i dati sperimentali (Raman) e la rigorosità della **Chimica Quantistica (DFT)** in tempo reale.

Gruppo Ricercatore /	Affiliazione	Focus Principale
<b>Jörg Behler</b>	Ruhr-Universität Bochum	Pioniere nello sviluppo dei <b>Neural Network Potentials (NNP)</b> , ovvero modelli AI addestrati sui calcoli DFT per eseguire simulazioni di meccanica quantistica con la velocità del <i>Machine Learning</i> . Questo è il fondamento per l'analisi in tempo reale.
<b>Matthias Scheffler</b>	Fritz Haber Institute of the Max Planck Society	Lavoro sulla gestione e l'utilizzo di enormi database di proprietà quantistiche ( <i>Nodal data</i> ), essenziale per addestrare i <b>Transformer Models</b> a mappare rapidamente gli spettri alle proprietà DFT.
<b>Mark E. Tuckerman</b>	New York University	Ricerca sulle simulazioni accelerate e sull'uso di metodi ibridi QM/MM (Quantistica/Meccanica Classica), che sono il prerequisito per l'integrazione AI-quantistica.

In sostanza, per l'**RL-Fase** guarda ai gruppi di *Autonomous Experimentation*; per la **Quantistica-AI** ai gruppi di *Neural Network Potentials*; e per la **Sintesi Spettrale** ai gruppi di *Generative Models for Chemistry*.

## Bibliografia: AI e Spettroscopia Raman

### I. Contributo Metodologico e Lavori degli Autori

Questa sezione include i riferimenti che hanno introdotto il concetto di pseudo-spettri Raman e la metodologia di base discussa.

1. **Sparavigna, A. C., & Gemini (Modello Linguistico di Google).** (2025). Il Ruolo del Reinforcement Learning nella Spettroscopia Raman. *Manoscritto di Lavoro. Manoscritto di Lavoro.*
2. **Sparavigna, A. C.** (2025). L'Intelligenza Artificiale e la Spettroscopia dei Minerali: Il Gesso come Ponte tra Analisi Storiche e Dati Moderni. *Manoscritto di Lavoro.*
3. **Sparavigna, A. C.** (2025). Rivelare l'Acqua di Cristallizzazione: Come l'AI Unisce Spettri Raman del Gesso da Tecniche d'Epoca e Database RRUFF. *Manoscritto di Lavoro.*

## II. Sintesi Spettroscopica Ipotetica (SS-I)

*(Generazione di spettri Raman/vibrazionali da strutture chimiche teoriche)*

4. **Cheng, A., Lo, A., Lee, K. L. K., Miret, S., & Aspuru-Guzik, A.** (2025). Stiefel Flow Matching for Moment-Constrained Structure Elucidation. *arXiv:2412.12540*. (Lavoro sull'uso di modelli generativi per l'elucidazione strutturale, guidato dal gruppo di Alán Aspuru-Guzik).
5. **Gastegger, M., Schütt, K. T., Schwaller, P., & Müller, K. R.** (2021). Machine learning of solvent effects on molecular spectra and reactions. *Chemical Science*, 12(34), 11394-11403. DOI: 10.1039/D1SC02742E. (Lavoro del gruppo di Klaus-Robert Müller sull'uso di modelli *surrogate* per simulare effetti che modificano gli spettri molecolari).

## III. Reinforcement Learning per la Scoperta Autonoma di Nuove Fasi (RL-Fase)

*(Sperimentazione autonoma con Ricompensa basata sulla "Novità Chimica")*

6. **Zhang, S., Song, Z., Godaliyadda, G. M. D. P., et al., & Simpson, G. J.** (2018). Dynamic sparse sampling for confocal Raman microscopy. *Analytical Chemistry*, 90(7), 4461-4469. DOI: 10.1021/acs.analchem.7b04749. (Lavoro fondamentale del gruppo di Gary J. Simpson sulla mappatura adattiva, il predecessore logico dell'RL in spettroscopia).
7. **Ede, J. M.** (2021). Adaptive partial scanning transmission electron microscopy with reinforcement learning. *Machine Learning: Science and Technology*, 2(4), 045011. DOI: 10.1088/2632-2153/abf5b6. (Un esempio metodologico diretto di RL per il campionamento adattivo basato sull'informazione in microscopia).
8. **Zou, Y., Cheng, A. H., Aldossary, A., et al., & Aspuru-Guzik, A.** (2025). El Agente: An autonomous agent for quantum chemistry. *Matter*, 8(7), 102263. DOI: 10.1016/j.matt.2025.102263. (Riferimento chiave sull'uso di agenti autonomi basati su LLM per flussi di lavoro complessi, come quelli guidati dal gruppo di Garnet Kin-Lic Chan).

## IV. Spettroscopia Quantistica-AI in Tempo Reale

*(Integrazione AI e Chimica Quantistica per l'analisi immediata delle proprietà)*

9. **Behler, J.** (2021). Four Generations of High-Dimensional Neural Network Potentials. *Chemical Reviews*, 121(16), 10037-10072. DOI: 10.1021/acs.chemrev.0c00868. (Panoramica essenziale di Jörg Behler sui *Neural Network Potentials*, che sono la tecnologia di base per l'accelerazione dei calcoli quantistici).
10. **Nair, A. S., Foppa, L., & Scheffler, M.** (2025). Materials Database from All-electron Hybrid Functional DFT Calculations. *Scientific Data*, 12, 1518. (Lavoro del gruppo di Matthias Scheffler che mostra come i database di calcoli rigorosi DFT siano fondamentali per addestrare l'AI alla mappatura delle proprietà).

11. **Chen, M. S., Robledo, A., Schäfer, C., et al., & Tuckerman, M. E.** (2025). Machine learning-accelerated path integral molecular dynamics simulations of reactive organic electrolytes. *The Journal of Chemical Physics*, 163, 144110. DOI: 10.1063/5.0208170. (Esempio del lavoro di Mark E. Tuckerman sull'accelerazione delle simulazioni di dinamica molecolare con il *Machine Learning*, essenziale per i modelli QM/MM in tempo reale).

## Conclusioni

I tre concetti di ricerca proposti—la Sintesi Spettroscopica Ipotetica, il Reinforcement Learning per la Scoperta di Nuove Fasi e la Spettroscopia Quantistica-AI in Tempo Reale—segnano il passaggio decisivo dell'AI dalla semplice analisi dei dati scientifici alla **guida attiva e autonoma del processo di scoperta**.

La ricerca futura dovrà concentrarsi sull'ingegneria dei sistemi complessi, in particolare:

1. **Sviluppo di Modelli Generativi Robusti:** Per garantire che gli spettri teorici generati dall'AI (SS-I) siano non solo plausibili, ma altamente predittivi e *filtrabili* per una sintesi mirata.
2. **Implementazione Operativa dell'RL per la Novità:** Trasformare il concetto di Ricompensa per la "Novità Chimica" in algoritmi RL robusti e in tempo reale, capaci di gestire le incertezze e il rumore degli ambienti sperimentali.
3. **Integrazione Ibrida dei Dati:** Creare architetture (come i Transformer) capaci di fondere la spettroscopia Raman con i dati di processo (*In Operando*) e i parametri quantistici per una diagnostica immediata.

L'impatto finale di questa convergenza è la trasformazione del tradizionale laboratorio di spettroscopia in un ecosistema di **sperimentazione autonoma**, dove l'AI agisce come un co-progettista scientifico in grado di ipotizzare, testare e validare in modo iterativo, accelerando in maniera esponenziale la scoperta di nuovi materiali e la comprensione dei fenomeni chimico-fisici.