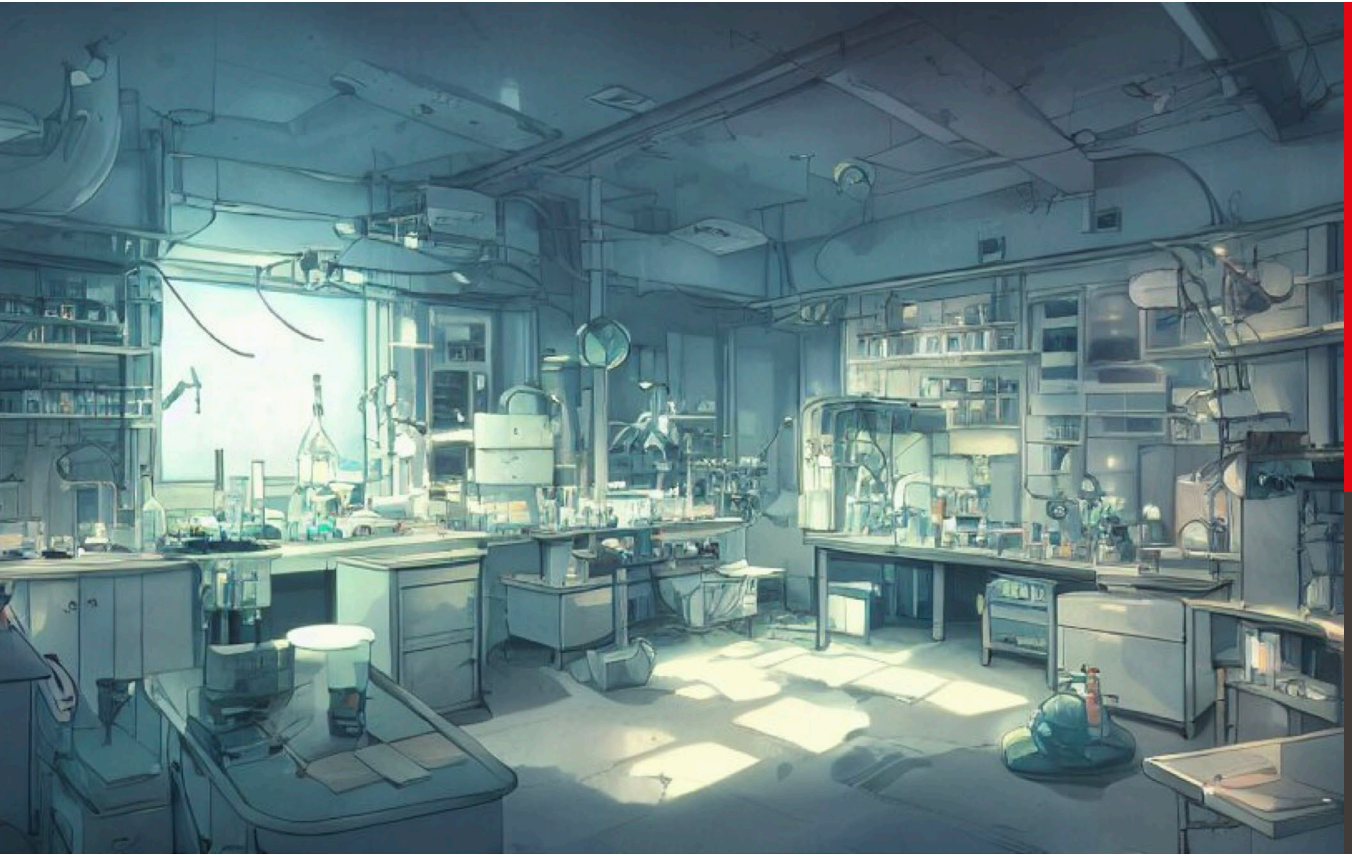
EPFL

NCCR Catalysis

LPDC

EPFL LIAC



The growing importance of AI in Chemistry: a brief overview

Victor Sabanza Gil

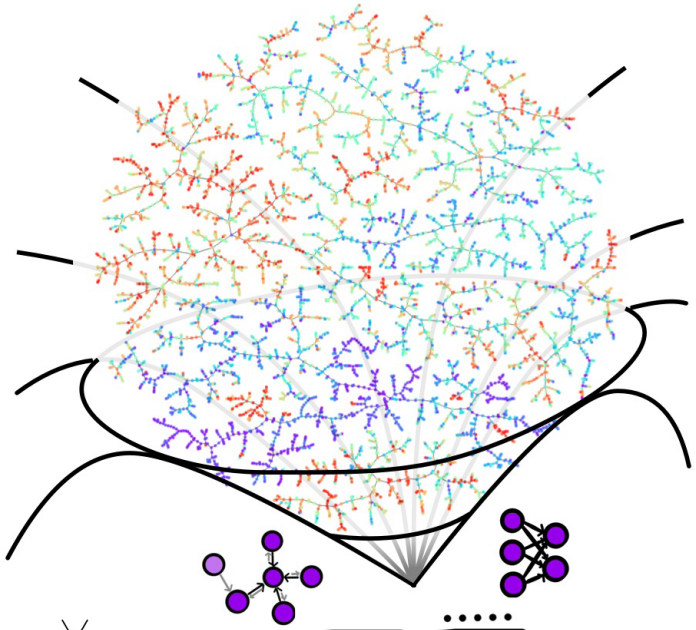
22/11/2023

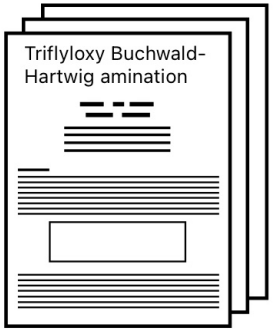
LIAC (Laboratoire d'Intelligence Artificielle Chimique)

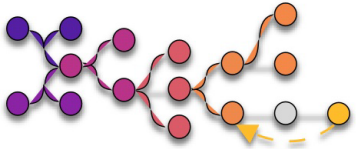
Victor Sabanza Gil

US20030166932A1: General Procedure H A solution of trifluoromethanesulfonic acid 3,5,8,8-tetramethyl-7,8-dihydronaphthalen-2-yl ester (Compound 35, 0.41 g, 1.2 mmol), Pd(OAc)2 (0.027 g, 0.12 mmol), BINAP (0.11 g, 0.18 mmol), Cs2CO3 (0.56 g, 1.72 mmol), ethyl 4-aminobenzoate (0.25 g, 1.5 mmol) and 5 mL of toluene was flushed with argon for 10 min, then stirred at 100° C. in a sealed tube for 48 h. ...

$$\hat{H}\Psi = E\Psi$$







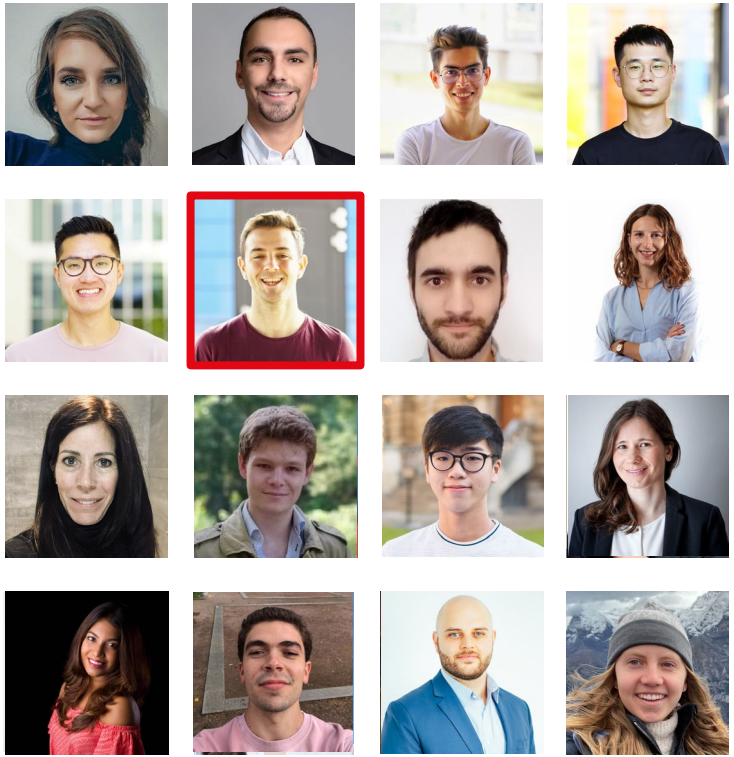
- Using AI to explore different problems and accelerate chemical discovery

EPFL

LIAC Team

PhD students
Bojana Ranković
Oliver Schilter (IBM Research)
Andres CM Bran
Junwu Chen
Jeff Guo
Victor Sabanza Gil (Luterbacher)
Paulo Neves (Janssen)
Rebecca Neeser (Correia, Vant)
Sarina Kopf (Nevado)
Daniel Armstrong
Joshua Sin (Roche)

Prof. Philippe Schwaller



>10 nationalities — one team!
https://schwallergroup.github.io

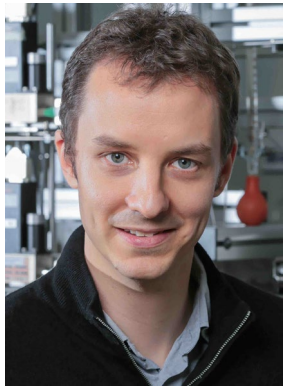
Admin
Annick Delmonaco

Postdocs
Geemi Wellawatte
Zlatko Jončev
Marta Brucka

Project students
Cheng-Hua Huang
Sacha Raffaud
Jenestin Anthonipillai
Théo Neukomm
Remi Schlama
Ziad El Maki

NCCR Catalysis
Swiss National Science Foundation
IBM Research
janssen
VANTAI
Roche

LPDC (Laboratoire de Procédés Durables et Catalytiques)



Prof. Jeremy Luterbacher

Overview

- AI in Chemistry
- 2 practical examples
- Future trends
- Conclusion

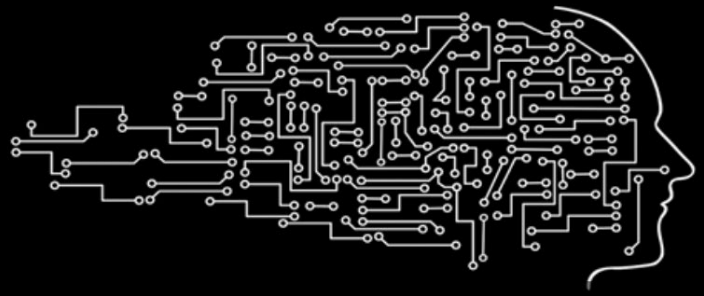
Artificial Intelligence in Chemistry

Why AI?

fundéu RAE

[buscador urgente de dudas]

RECOMENDACIONES CONSULTAS CATEGORÍAS ESPECIALES BLOG



29/12/2022

inteligencia artificial es la expresión del 2022 para la FundéuRAE

Escuchar

La Fundación del Español Urgente (FundéuRAE), promovida por la Agencia EFE y la Real Academia Española, otorga el título de palabra del año a la expresión compleja *inteligencia artificial*.

Esta construcción está definida en el diccionario académico como ‘disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico’.

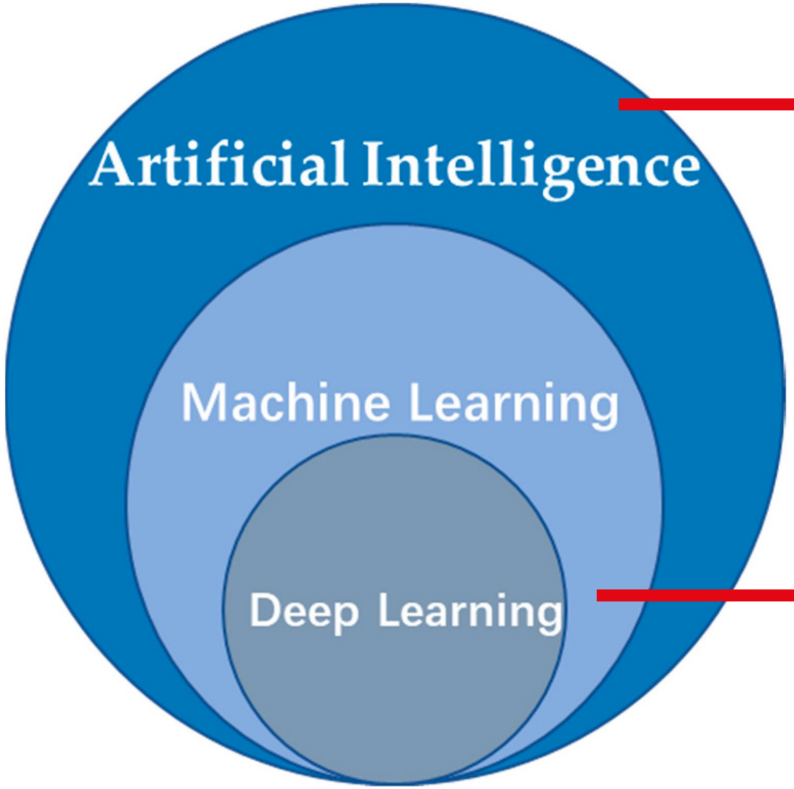
Este concepto se incorporó al diccionario de la Academia en su edición de 1992, y este año la FundéuRAE lo ha seleccionado por su importante presencia en los medios de comunicación durante estos últimos doce meses, así como en el debate social, debido a los diversos avances desarrollados en este ámbito y las consecuencias éticas derivadas.

”Artificial Intelligence is the 2022 expression”

REAL ACADEMIA ESPAÑOLA

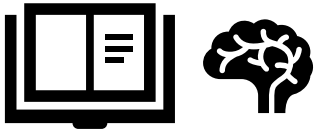
What is Artificial Intelligence?

Victor Sabanza Gil

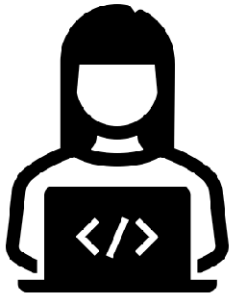


Computer mimics human intelligence

Program learns without previous rules



Traditional programming (“Expert system”)



Input x

Output y

5.5

If grade 4 or higher (x >= 4), student passes the course.

Pass

3.5

Fail

Predefined human-written rules (knowledge base)

Supervised Machine Learning

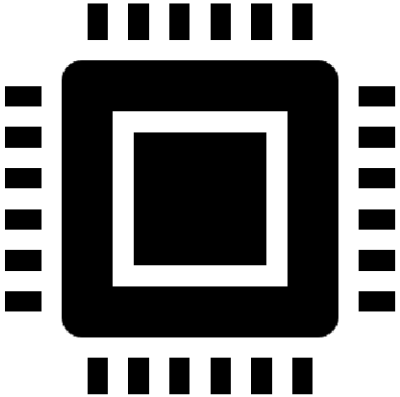
Victor Sabanza Gil

Input x

Model

Output y

Data (training + validation)



dog

cat

cat

Unseen examples (test)

dog

And why in chemistry?

How can AI accelerate chemical research?

Practical examples

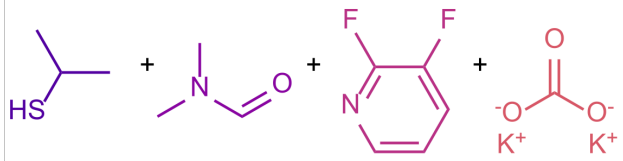
Speaker

Reaction prediction

- Turn it into a translation task!

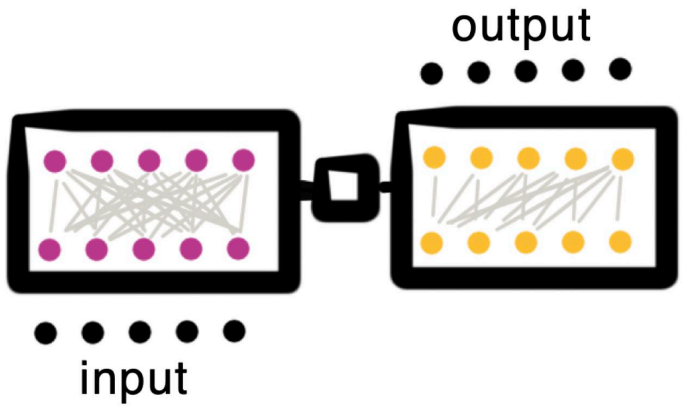
El gato es negro

precursors

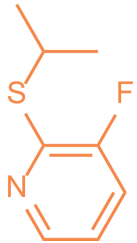


CC(C)S.CN(C)C=O.Fc1cccnc1F.

Le chat est noir



products



CC(C)Sc1ncccc1F

~90% accuracy!

ACS Cent. Sci., *5*, **9**, 1572–1583 (2019)

Here comes the robot!

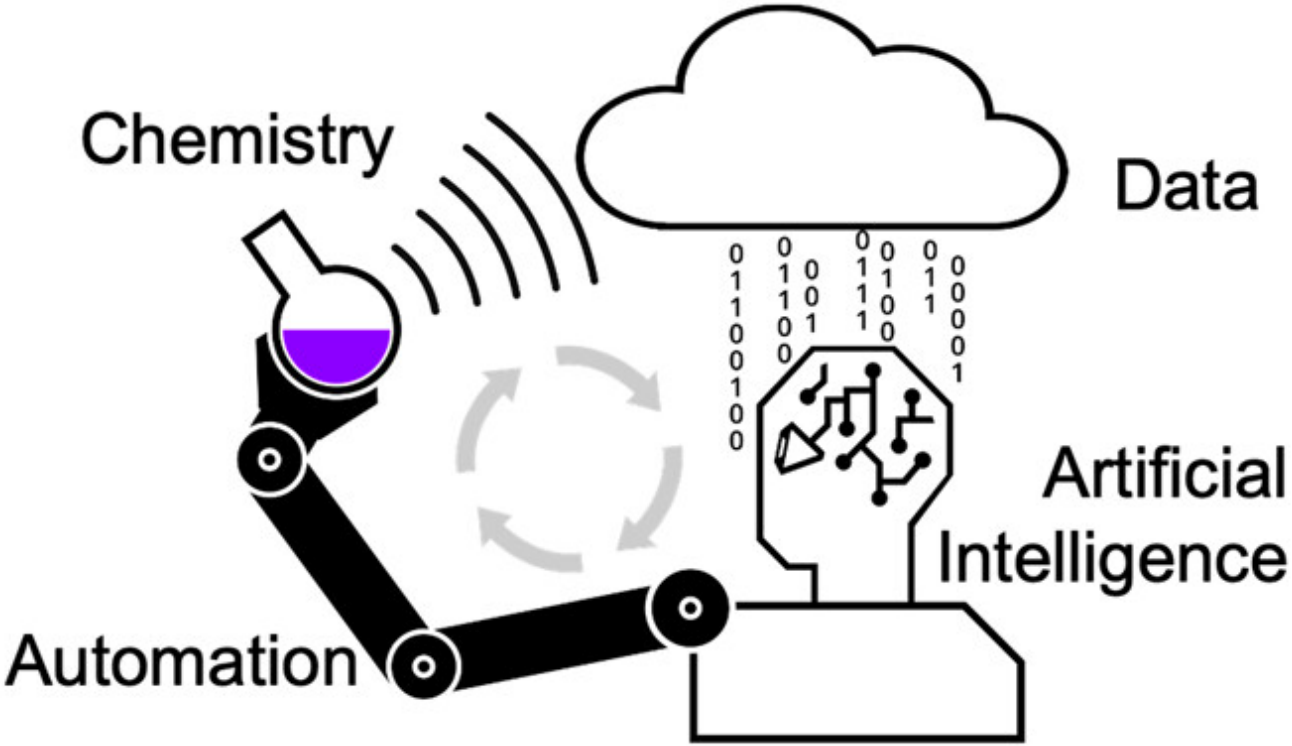
Nature **583**, 237–241 (2020)

- Robots can do experiments!

Future trends

Victor Sabanza Gil

Self-driving labs: the closed-loop paradigm



Acc. Chem. Res., **55**, 17, 2454–2466 (2022)

Are (Large) Language Models the future of chemistry?

nature reviews chemistry

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature reviews chemistry > comment > article

Comment | Published: 19 May 2023

The future of chemistry is language

Andrew D. White

Nature Reviews Chemistry **7**, 457–458 (2023) | Cite this article

5211 Accesses | **2** Citations | **92** Altmetric | Metrics

Large language models such as GPT-4 have been approaching human-level ability across many expert domains. GPT-4 can accomplish complex tasks in chemistry purely from English instructions, which may transform the future of chemistry.

Large language models (LLMs) predict an output sequence from an input sequence. For example, you could input "ethanol" and it will output CCO – the simplified molecular-input line-entry system (SMILES) representation of ethanol (SMILES are how a chemical structure is written as text). Like any machine learning model, LLMs are fit empirically with a large dataset – generally a large subset of the internet. While they were first pursued for tasks in natural language, such as translating English to French[1], they can now be used to identify objects in images[2], predict protein structure[3] and estimate reaction yields[4], and they are the technology behind the popular ChatGPT.

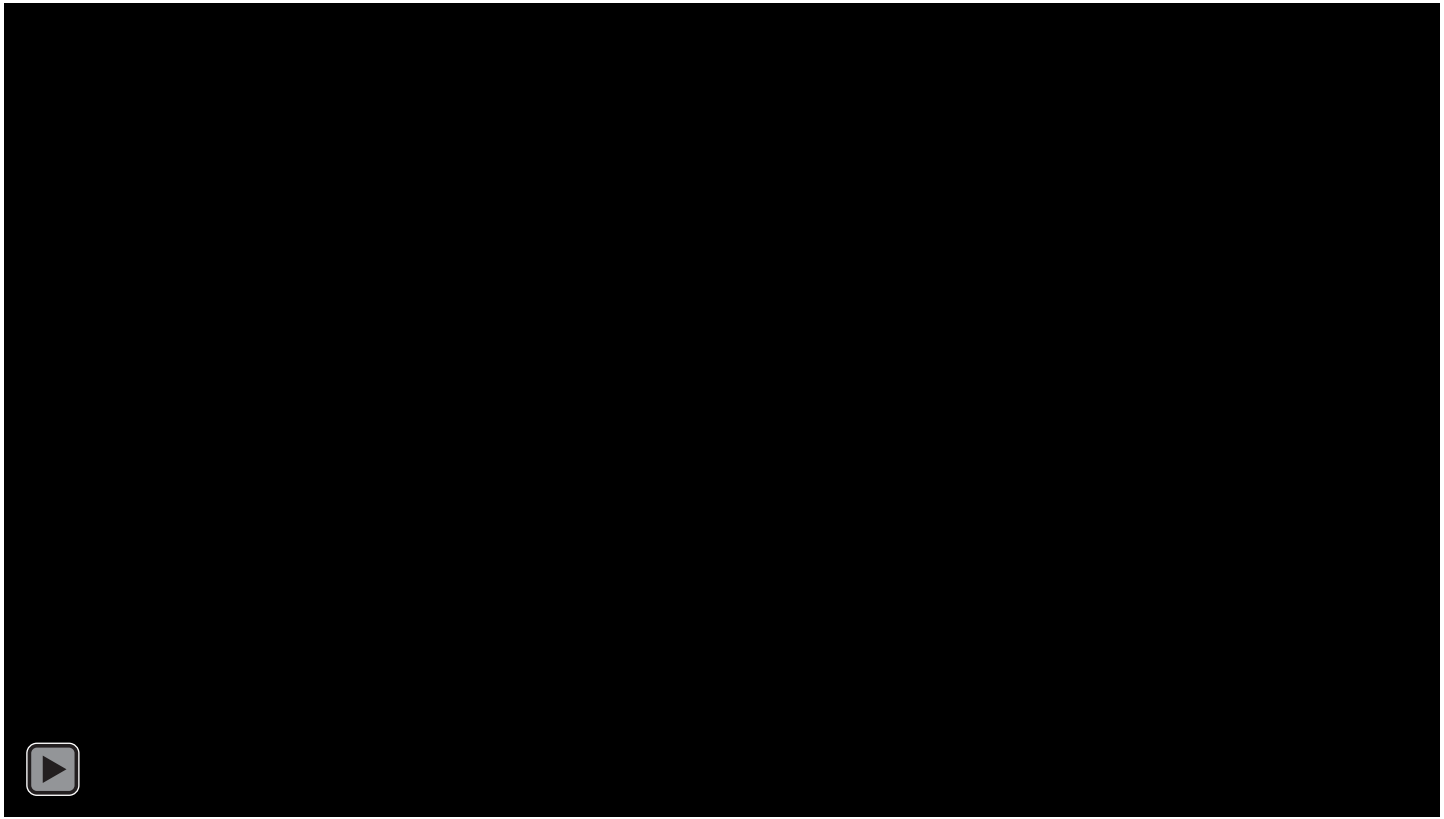
Bard

Meta

+

Nat Rev Chem **7**, 457–458 (2023)

ChemCrow: AI can use now chemistry tools



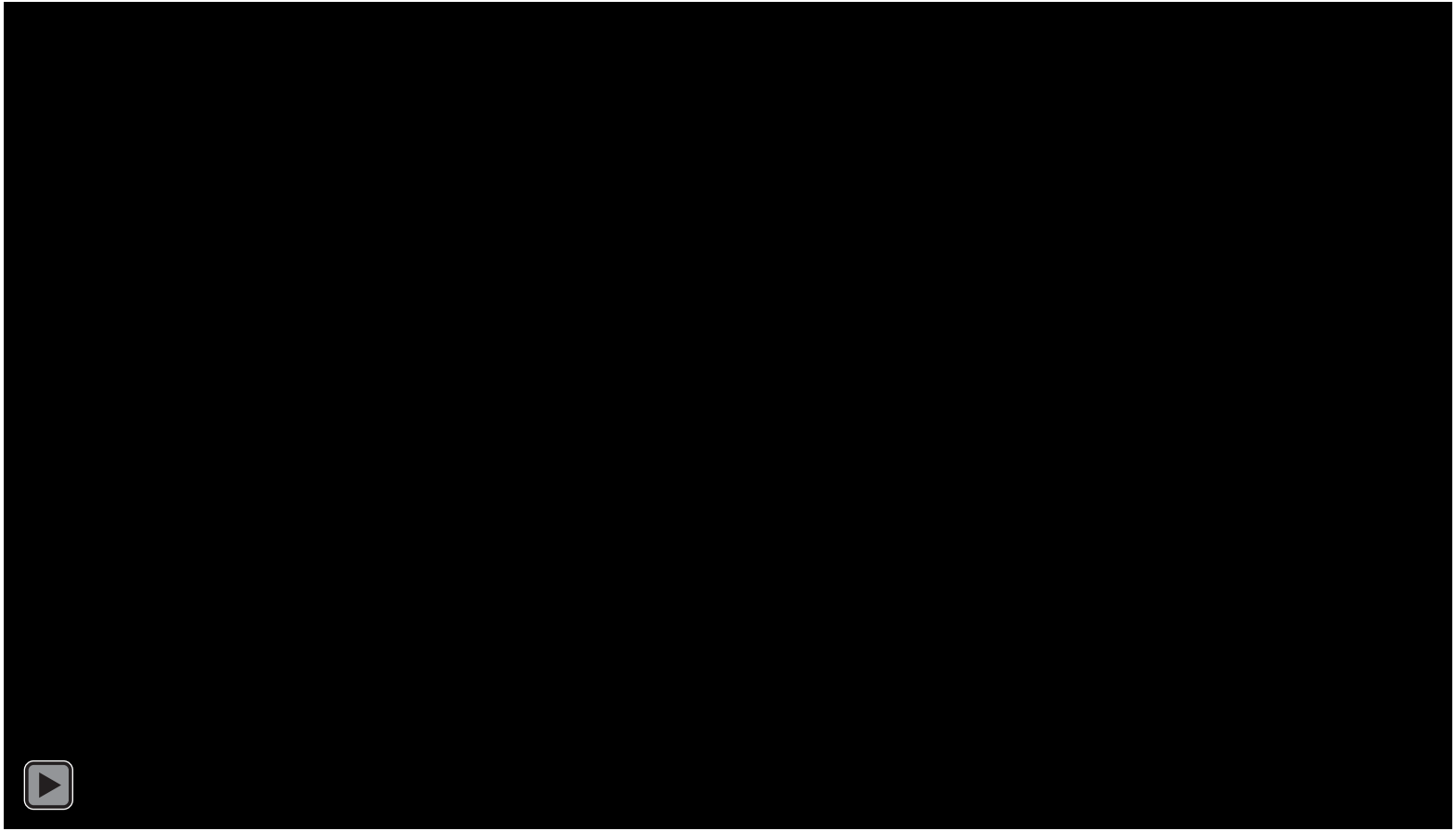
https://huggingface.co/spaces/doncamilom/ChemCrow

Reimagining chemical education

Victor Sabanza Gil

- I would like to design an experiment to teach high school students how to synthesize aspirin. Write an experimental protocol indicating the substances, conditions, and method that we will use, and comment on the most important details to consider in order to maximize the learning experience.

Reimagining chemical education

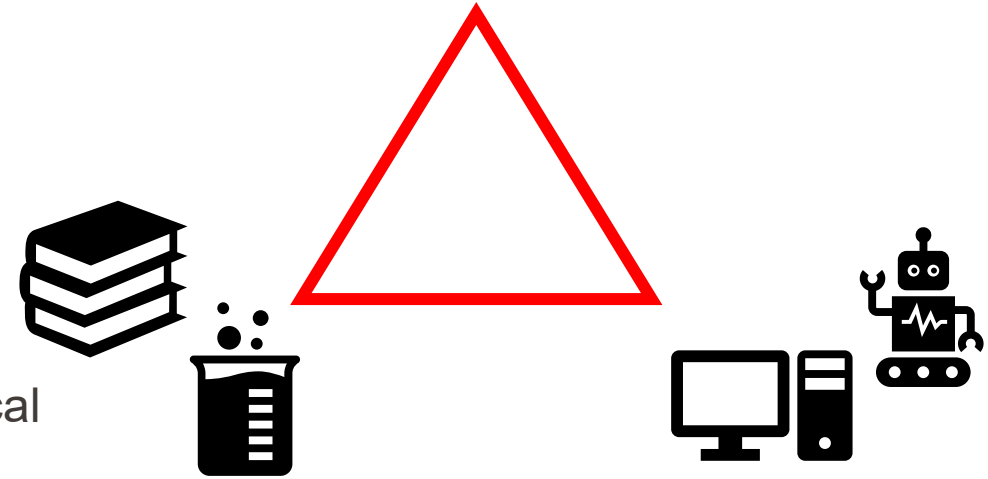


Conclusion

Victor Sabanza Gil

What will future chemists need?

Creativity

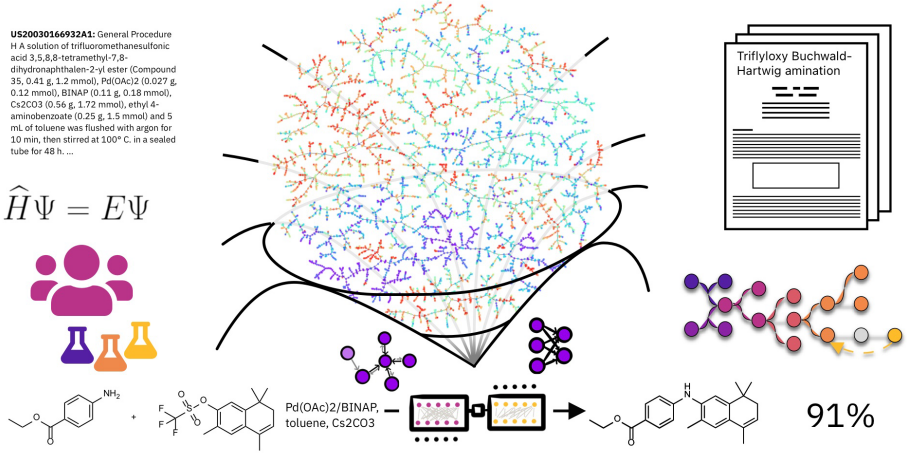


Solid chemical knowledge

Digital skills (AI & automation)

Towards the chemist of the future

Thanks for your attention!



victor.sabanzagil@epfl.ch

@VictorSabanza
@SchwallerGroup
@LPDC_EPFL

victorsabanza
schwallergroup