

An aerial photograph of the EPFL campus in Lausanne, Switzerland, serves as the background. It shows modern university buildings, a large lake (Lac de lausane), and distant mountains under a dramatic, cloudy sky. A large red rectangular box is overlaid on the right side of the image, containing the title text.

# Outils de recherche d'information en chimie : exemples + tendances

**Alain Borel**

EPFL - Bibliothèque  
[alain.borel@epfl.ch](mailto:alain.borel@epfl.ch)

9 mai 2025

# A propos de ce document



Ce document est distribué sous licence  
Creative Commons - Attribution

**CC-BY**

# Buts

Vous familiariser avec SciFinder, Reaxys, Chemspider, PubChem:

- Contenus
- Droits
- Principes de recherche d'information
- Perspectives (IA...)

Dépendant de l'intérêt:

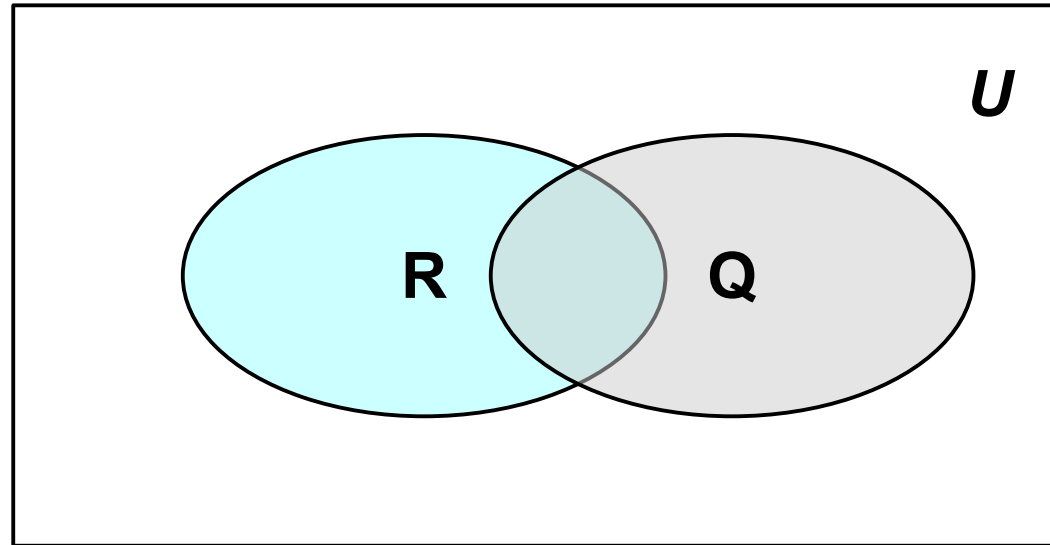
- Gestion de bibliographie
- Finir avant 18h?

# SciFinder: enregistrement obligatoire

Créer un compte avec une adresse EPFL sur  
<https://scifinder.cas.org>

En cas d'accès hors de l'EPFL (y.c. 4G/5G): VPN

# Recherche d'information : concepts de base



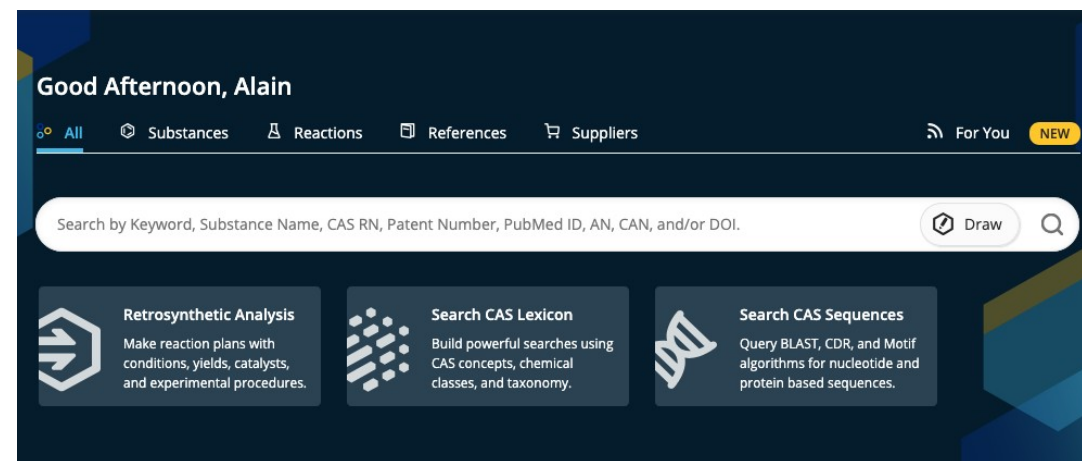
U: Universe

R: records you will consider Relevant

Q: records found by your Query



# SciFinder



CAPLus: > 59 mio de référence (livres, conférences, brevets); 1'500 journaux complets dès 1907 (50'000 titres plus ou moins couverts), 180'000 références avant 1907. (+1.5 mio de 2022 à 2023, pas de nouveaux chiffres 2024 ou 2025)

ChemZent: références historiques 1830-1969 (traduction en anglais)

Registry: > 219 mio de composés (+15 mio entre 2023 et 2024, pas de nouveaux chiffres en 2025)

CASREACT: > 150 mio de réactions (pas de nouveaux chiffres en 2025)

Catalogues de produits

MEDLINE: 5200 journaux dès 1950, > 33 mio de références

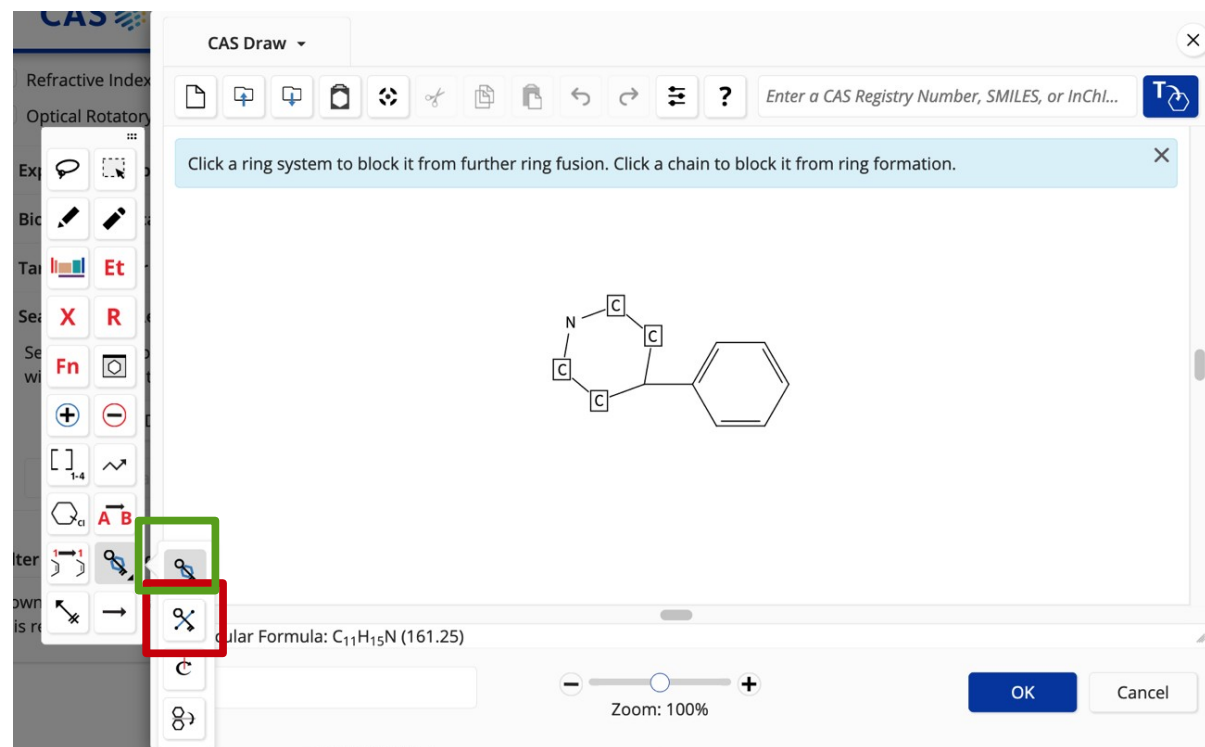
# Recherche de structures: différents types

- Structure exacte
- Sous-structure: le fragment fait partie d'une structure plus grande
- Similarité: autorise le remplacement d'atomes ou de liaisons
- Structures de Markush: contiennent des groupes génériques (-R, -X, etc.) ; la recherche trouve des brevets au lieu de composés

**=> principe: recherche d'un type d'information (publication, structure ou réaction) supportée par des filtres**

# Trop de structures dans Scifinder?

- Utiliser des fragments plus détaillés
- H explicites ou **lock atoms** pour éviter les substituants indésirables
- **Lock rings**
- Utiliser les filtres (substance role, Refine, etc.)



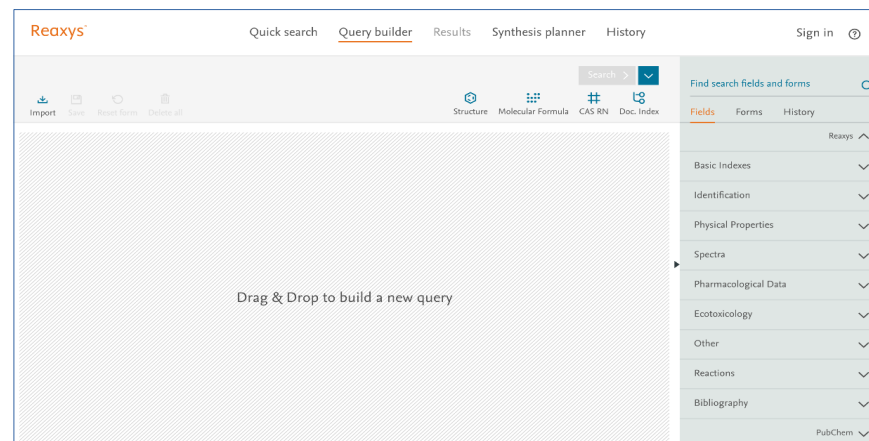


# Reaxys

Beilstein: chimie organique (175 journaux; dès 1771), 8'000'000 composés.



Gmelin: chimie minérale (62 journaux; dès 1772), 2'200'000 composés



+ nouveaux contenus (Patent Chemistry database, PubChem...)  
=> 283 mio de composés (+23 mio), > 73 mio de réactions

# Reaxys: enregistrement facultatif

Certaines fonctionnalités nécessitent un compte personnel

Accès hors de l'EPFL (y.c. 4G/5G): VPN

# Recherches par propriétés dans Reaxys

Reaxys contient > 500 mio de faits expérimentaux  
=> **possibilité de recherches très pointues**

Composés contenant au moins un  
atome d'uranium

+

Etudié par des calculs quantiques

+

Etudié par RMN

+

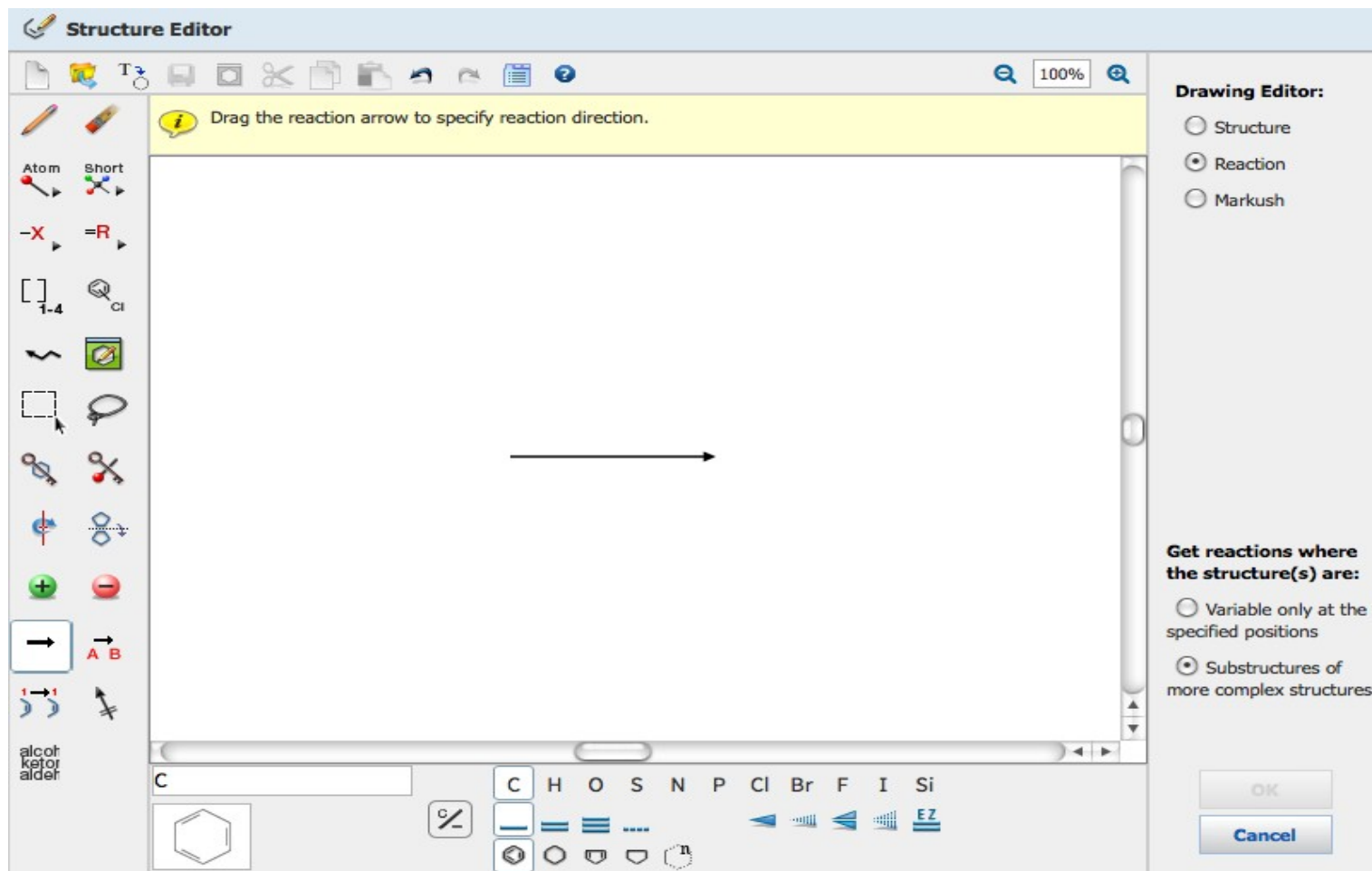
Point de fusion > 100 C

The screenshot shows the Reaxys Query Builder interface. The top navigation bar includes 'Quick search', 'Query builder' (selected), 'Results', 'Synthesis planner', and 'History'. Below the navigation bar, there are icons for 'Import', 'Save', 'Reset form', and 'Delete all'. On the right, there are icons for 'Structure', 'Molecular Formula', 'CAS RN', and 'Doc. Index'. The main query builder area displays a list of criteria:

- Quantum Chemical Calculations (Find any, Show fields)
- NMR Spectroscopy (Find any, Show fields)
- Melting Point (Find any, Hide fields)
- Structure (Find any, Hide fields)

The 'Melting Point' criterion is expanded, showing a dropdown menu with '>' selected, and a text input field containing '100'. Below this, there is a section for 'Solvent (Melting Point)' with a dropdown menu and a text input field. The 'Structure' criterion is also expanded, showing a large empty box for drawing the chemical structure. At the bottom of the structure box, it says 'As drawn'.

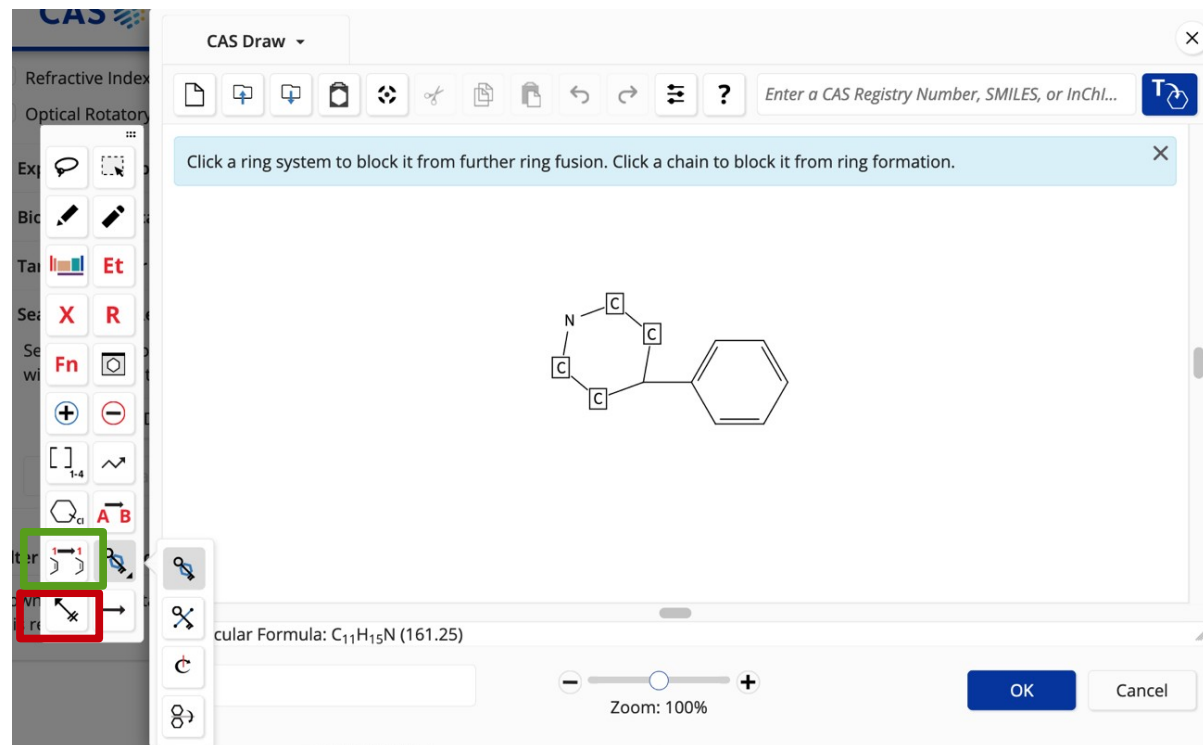
# Recherche de réactions



A priori simple: flèche de réaction entre les réactants et les produits

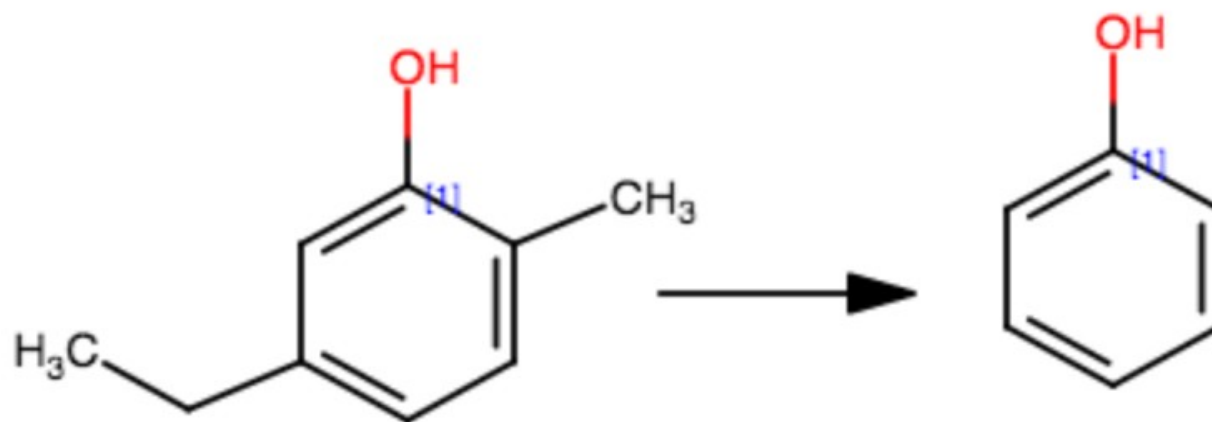
# Trop de réactions dans Scifinder?

abandonné



- **Map atoms**: associe les atomes des reactants avec ceux des produits
- **Mark bonds**: spécifie les liaisons des réactants qui seront rompues ou celles des produits qui seront formées
- **Functional groups**: ignore +/- la structure, recherche des réactions qui transforment un groupe fonctionnel choisi dans une liste en un autre

# Recherche de réactions dans Reaxys



Recherches de structures et de réactions similaires à Scifinder, mais menus différents



# Combinaisons de résultats : Scifinder-n

# Propriétés: autres sources possibles

- CRC Handbook of Chemistry and Physics
- NIST Chemistry Webbook  
<http://webbook.nist.gov/chemistry/>
- Springer Materials <http://www.springermaterials.com>
- ...

> 120 mio de structures

Aggrégation de sources variées (catalogues, bases chiminformatiques + input communautaire, cf. <https://www.chemspider.com/Datasources> )

Projet indépendant repris par la Royal Society of Chemistry; beaucoup plus ouvert (UI + RSC Compounds API)

Accès libre et gratuit

# PubChem

~ 119 mio de structures (+1 mio)

Développé par la National Library of Medicine (NIH, USA), aggrégation de sources (catalogues, éditeurs scientifiques, agences gouvernementales... )

Liens avec les publications (PubMed) fournis par des tiers (ex. IBM):  
Kim, S.; Thiessen, P. A.; Cheng, T.; Yu, B.; Shoemaker, B. A.; Wang, J.; Bolton, E. E.; Wang, Y.; Bryant, S. H. Literature Information in PubChem: Associations between PubChem Records and Scientific Articles. J Cheminform 2016, 8, 32. <https://doi.org/10.1186/s13321-016-0142-6>.

Accès libre et gratuit

- Rétrosynthèse
- IA

# Tendances actuelles

# Rétrosynthèse

Evolution logique des bases de données de réaction: conception de plans de synthèse. SciFinder et Reaxys sont en compétition pour le développement de cette fonctionnalité. Chronologiquement :

- [2009] Reaxys Synthesis Planner (sélection de réactions par l'utilisateur)
- [2011] Scifinder SciPlanner (réactions et autres informations stockées dans un document général)
- [2020] Scifinder-n "Create Synthesis plan"
- [début 2022] Reaxys Predictive Retrosynthesis [pas disponible pour l'EPFL]
- [fin 2022] Scifinder-n Retrosynthesis Plan
- [2025?] Interactivité dans les plans de Scifinder

**=> tendance: aide à la décision, construction de résultats concrets**



# Et l'IA dans tout ça ?

**Développement des fonctionnalités, particulièrement pour la planification de synthèses** Exemple: Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. Nature 2018, 555 (7698), 604–610. <https://doi.org/10.1038/nature25978>. => Reaxys Predictive Retrosynthesis

**Recherche en langage naturel** facilitée <https://www.cas.org/solutions/cas-scifinder-discovery-platform/cas-scifinder/upcoming-enhancements>

## Nouvelle concurrence: ChatGPT et autres LLMs

Schulze Balhorn, L.; Weber, J. M.; Buijsman, S.; Hildebrandt, J. R.; Ziefle, M.; Schweidtmann, A. M. Empirical Assessment of ChatGPT's Answering Capabilities in Natural Science and Engineering. Sci Rep 2024, 14 (1), 4998. <https://doi.org/10.1038/s41598-024-54936-7>.

# Modèles/motifs/patrons/templates/patterns/idioms

En informatique, on traite souvent un problème à l'aide de structures réutilisables plus ou moins abstraites.

Pattern: arrangement caractéristique d'éléments reconnus comme une bonne pratique en réponse à un problème (FR: motif, modèle)

Template: modèle comprenant quelques éléments modifiables dans une structure réutilisable (FR: modèle, gabarit, patron)

Idiom: expression toute faite qu'on réutilise assez systématiquement dans un contexte donné (FR: idiome, idiotisme ;-)

# AI & Patterns

Conception de LLMs: <https://eugeneyan.com/writing/llm-patterns/>

Prompt patterns: <https://medium.com/@corraljrmiguel/21-prompt-patterns-you-should-know-636c931bba2a>

Méthodologies plus élaborées, facilitées par des templates

ReAct (Reason + Act): <https://www.promptingguide.ai/techniques/react>

Reflexion: <https://promptengineering.org/reflexion-an-iterative-approach-to-llm-problem-solving/>

AutoGPT: <https://medium.com/@Jstnwrds>

55/a-prompt-template-for-generating-autogpt-input-summaries-with-chatgpt-a98388059673

# Exemples

ReAct:

<https://claude.ai/share/b8a6c1fb-6803-4092-8766-e277e857cf16>

Reflexion:

<https://claude.ai/chat/d5595641-1979-4b75-964f-fc2f4b3375db>

# Votre avis nous intéresse



<https://library-survey.epfl.ch/c/unkgggkud>

[alain.borel@epfl.ch](mailto:alain.borel@epfl.ch)