

Aller plus loin avec OpenRefine

API, intelligence artificielle et outils avancés

Journée LibreABC, 9 septembre 2025, Genève

Thomas Guignard, TG Consulting, thomas@tgconsulting.ca

Lionel Walter, arbim IT, lionel.walter@arbim.ch



Présentations

Plan de l'atelier

- 10h15 – 11h Récupération des données depuis le Web dans OpenRefine (Lionel)
- 11h – 11h45 Utiliser de l'IA pour traiter les données dans OpenRefine (Thomas)

Prérequis et Support

- cf. [Github](#)
- Y compris cette présentation

Partie 1

Récupération de données depuis le Web

OpenRefine

- <https://openrefine.org/>
- Library Carpentry
<https://librarycarpentry.github.io/lc-open-refine/>

Analogie «Au café»

Rachèle

J'aimerais un
café avec du
lait d'avoine et
un cookie

Plus de lait
d'avoine.
Grand ou
petit ?



Barista

Un petit café
lait de soja et
un cookie pour
la table 37



Cuisine



Adapted from Joshua Dull, Sawyer Newman, & David Cirella. (2019, October). APIs for Libraries (Version v1.0). Zenodo. CC-BY. <http://doi.org/10.5281/zenodo.3493358>

Analogie «Au café»

Rachere

`https://moncafe.ch/order?
beverage=coffee&snack=cookie`

API

Application



Request

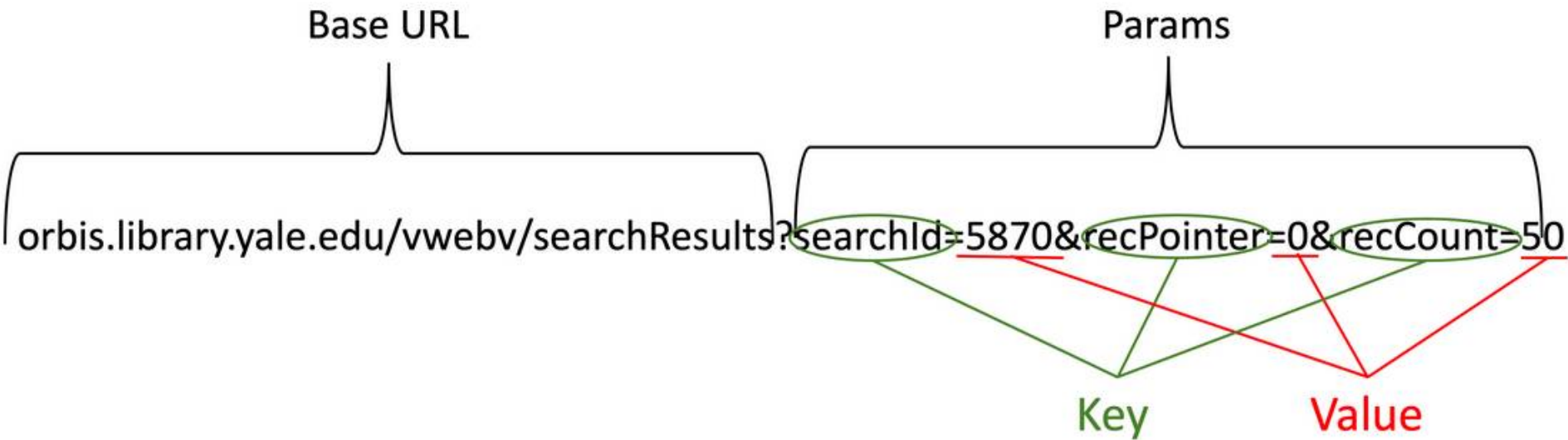
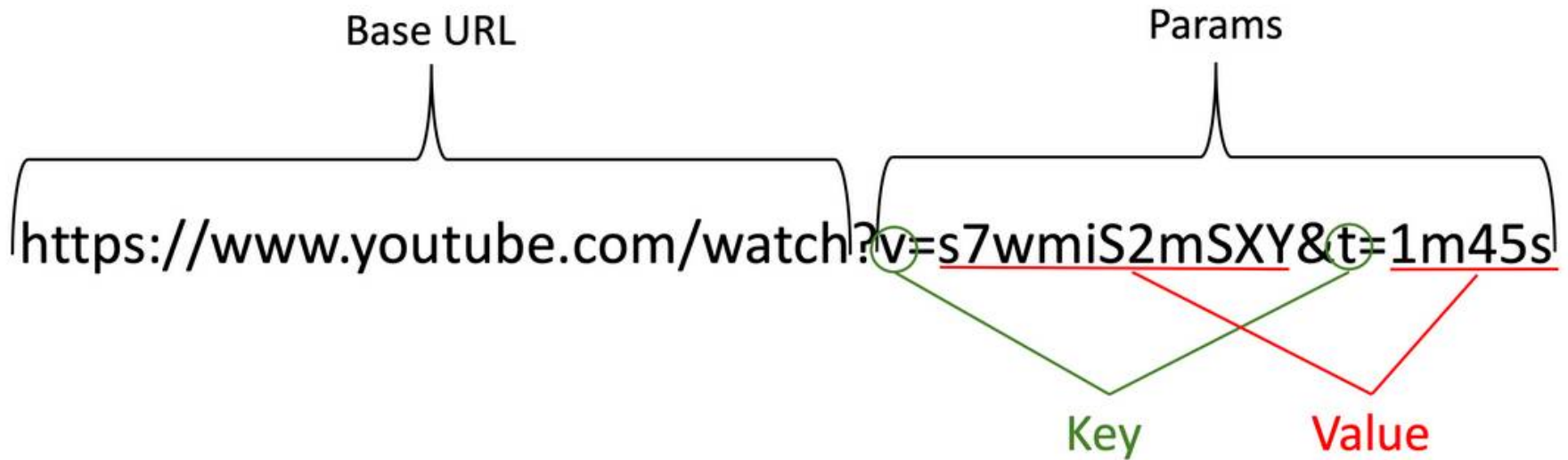
Response

```
{  
  "items": [  
    {  
      "type": "beverage",  
      "name": "coffee"  
    },  
    {  
      "type": "snack",  
      "name": "cookie"  
    }  
  ]  
}
```



Adapted from Joshua Dull, Sawyer Newman, & David Cirella. (2019, October). APIs for Libraries (Version v1.0). Zenodo. CC-BY. <http://doi.org/10.5281/zenodo.3493358>

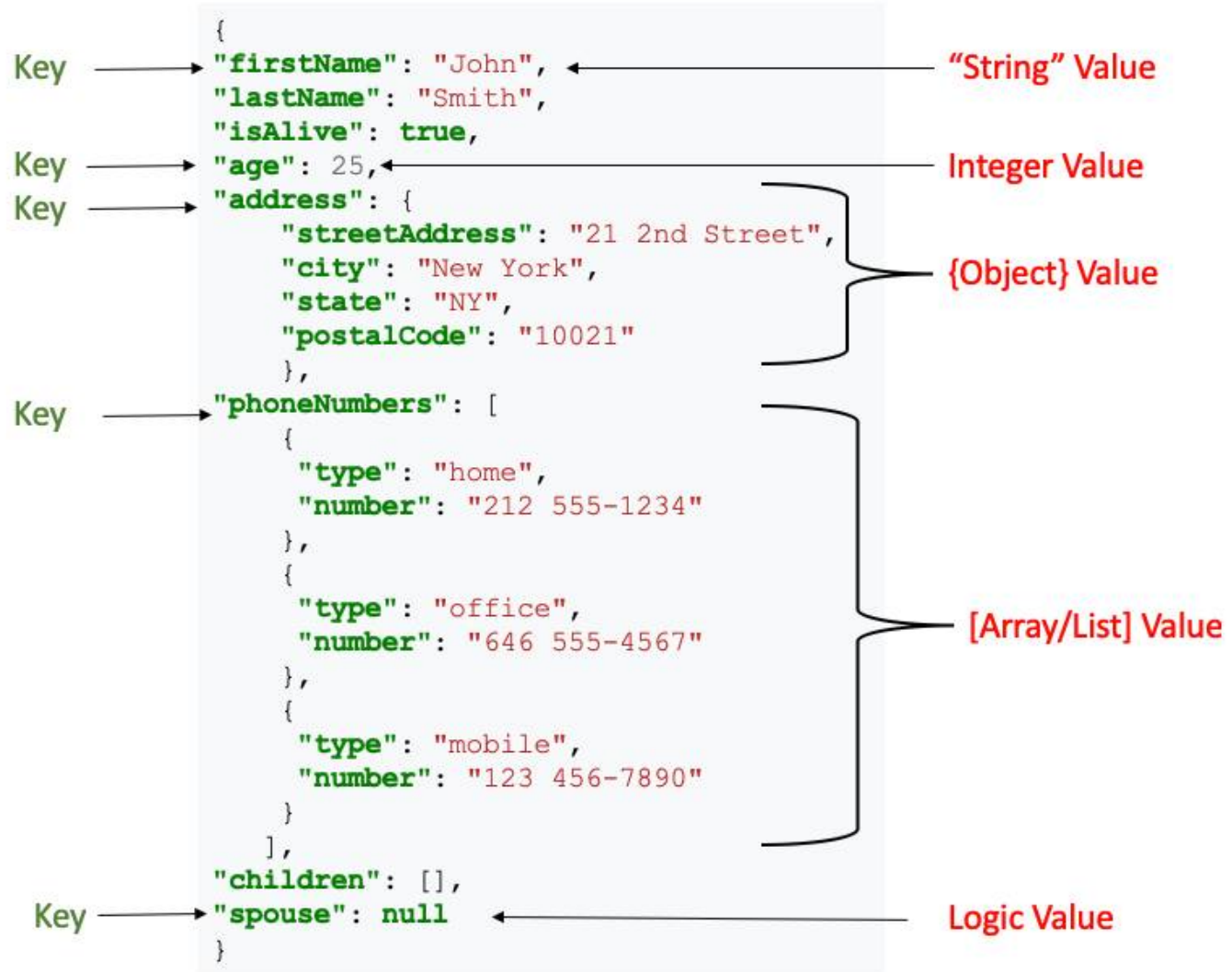
Structure d'URLs



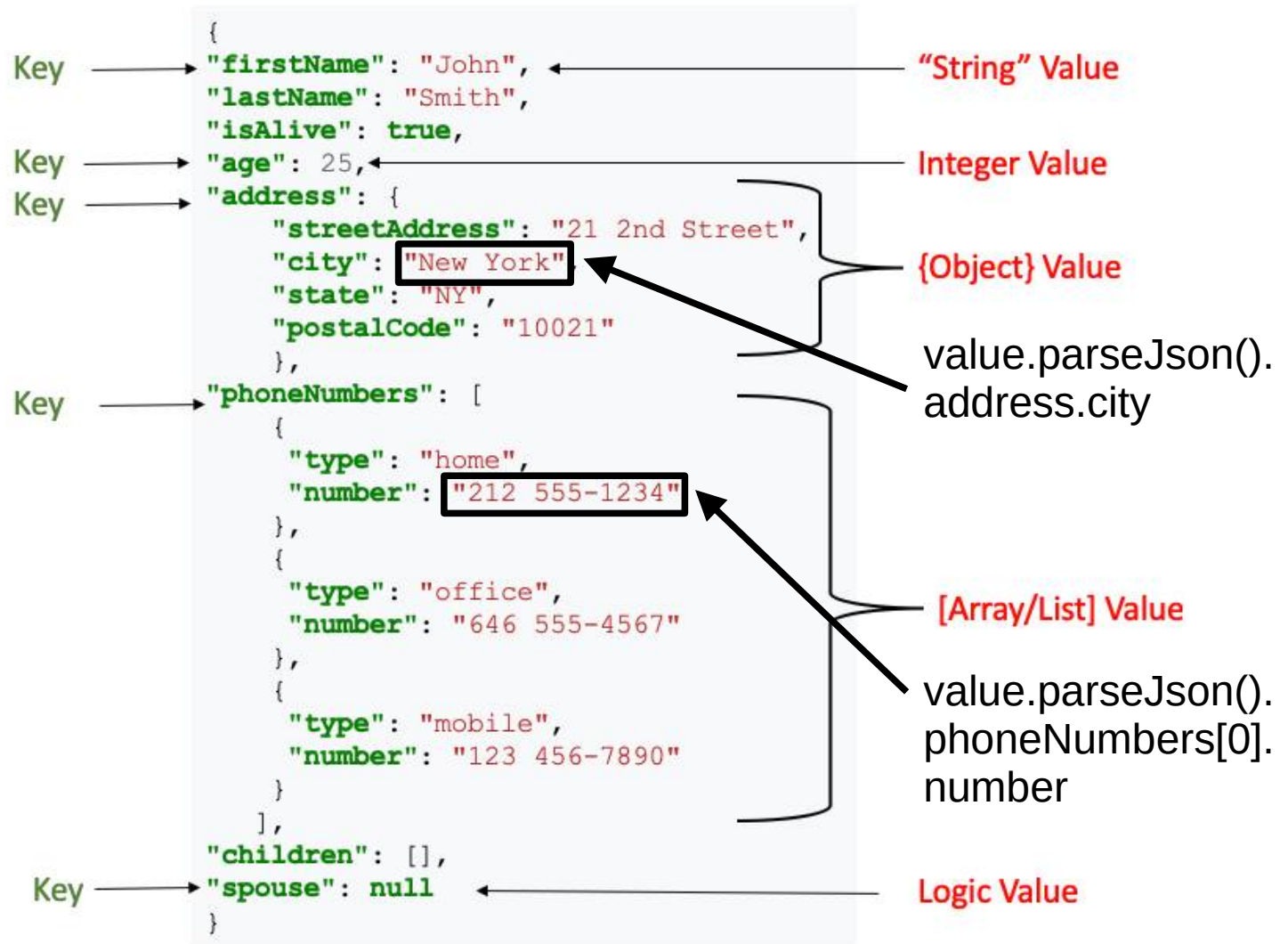
Assembler des valeurs dans OpenRefine

- `https://lobid.org/gnd/search?q=Maillart, Ella&format=json`
- `"https://lobid.org/gnd/search?q="`
- `+value+`
- `"&format=json"`
- Mieux : `+value.escape("url")+`

JSON



JSON in OpenRefine



3 méthodes dans OpenRefine

- Réconciliation
- Récupération de données via API (json)
- Récupération de données depuis le web directement (html)

Démo Réconciliation Wikidata

- Récupérer le lieu de naissance

Démo API GND

- Récupérer des informations sur Ella Maillart depuis la GND
- <https://lobid.org/gnd/119081288>
- Doc API : <https://lobid.org/gnd/api>
- <https://lobid.org/gnd/search?q=Maillart,%20Ella&filter=type%3APerson&format=json>

Exercices

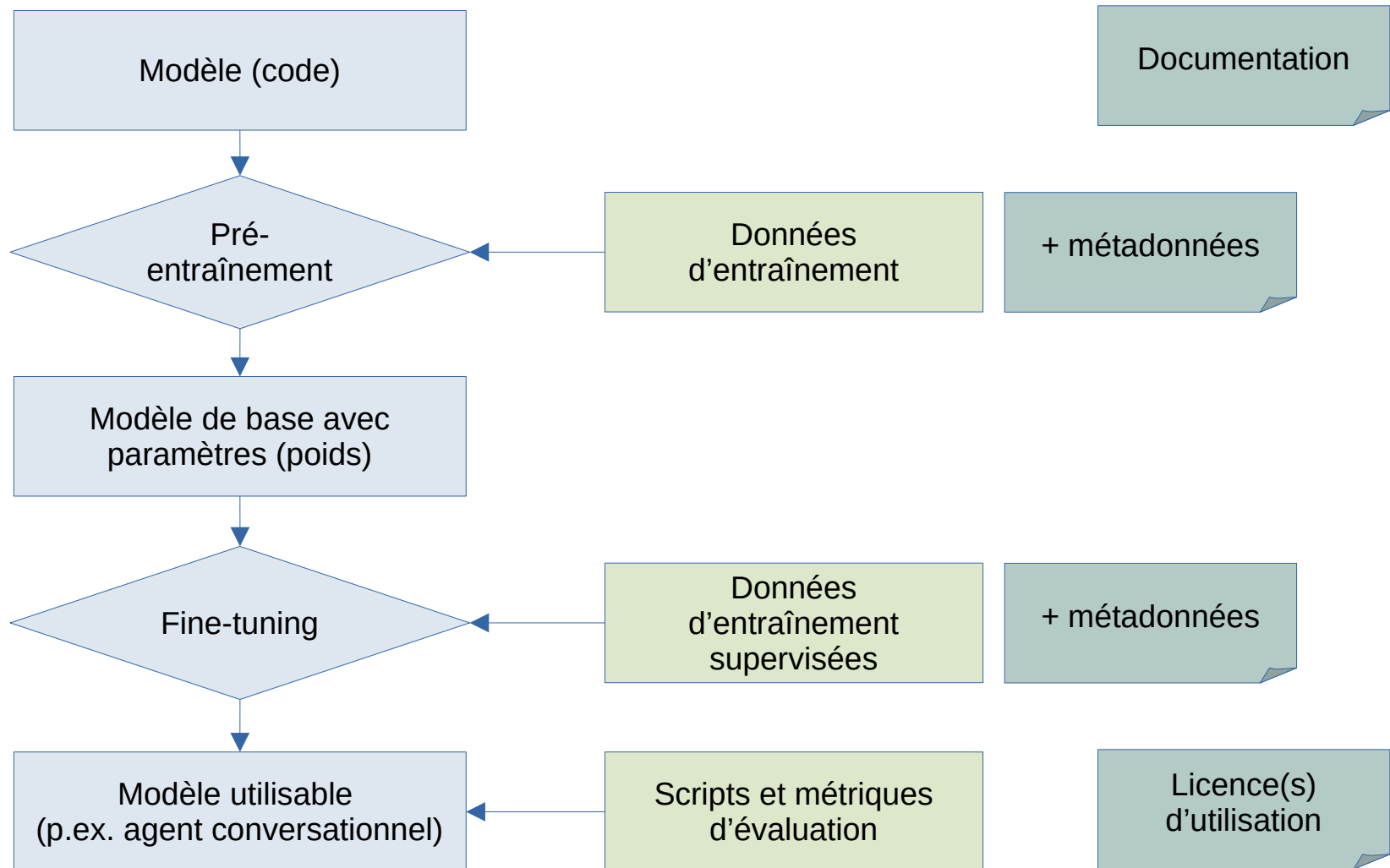
- Cf.

<https://github.com/liowalter/open-refine-libreabc2025/blob/main/Exercices.md>

Partie 2

Utiliser l'IA pour traiter les données dans OpenRefine

Les LLM en (très) bref

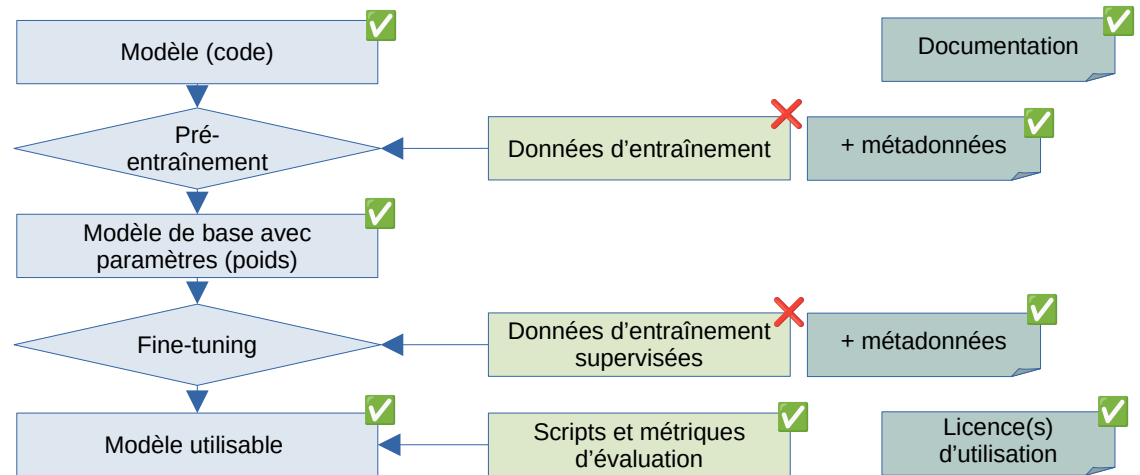


Ouvert ou pas ?



**open source
initiative®**

Open Source AI Definition
(OSAID 1.0)

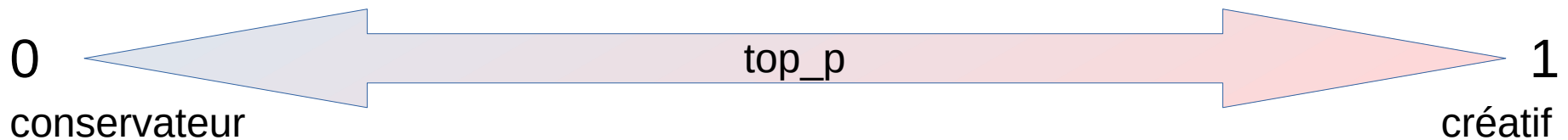
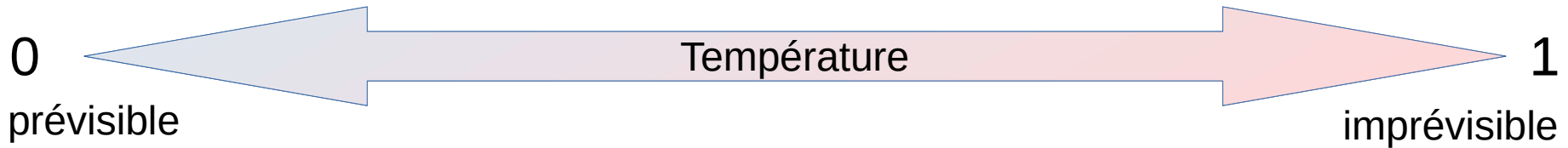


✓ Ouverts : Pythia (Eleuther AI), OLMo (AI2), Amber & CrystalCoder (LLM360), T5 (Google) ...

⚠ Semi-ouverts : Llama2 (Meta), Grok (X/Twitter), Phi-2 (Microsoft), Mixtral (Mistral) ...

✗ Fermés : GPT-4o (OpenAI), Gemini (Google), Claude (Anthropic) ...

Température et top_p



Passons à la pratique !

- Options à choix :
 - API Infomaniak avec la clé de l'atelier
 - API de votre choix avec votre propre clé
 - LLM en local avec ollama
 - Modèle suisse Apertus via API PublicAI

Sources

- Joshua Dull, Sawyer Newman, & David Cirella. (2019, October). APIs for Libraries (Version v1.0). Zenodo. CC-BY. <https://doi.org/10.5281/zenodo.3493358>
- Alammam, J., & Grootendorst, M. (2024). Hands-on large language models: Language understanding and generation (First edition). O'Reilly Media, Inc. <https://www.oreilly.com/library/view/-/9781098150952/>
- Open Source Alliance. (2025). Open Weight Definition (OWD). Retrieved Aug 6, 2025 from <https://openweight.org/>
- Open Source Initiative. (n.d.). The Open Source AI Definition (OSAID 1.0). Retrieved May 27, 2025, from <https://opensource.org/ai/open-source-ai-definition>
- Sowe, S., Mou, Y., Cheng, D., Kong, L., Neumann, A. T., & Decker, S. (2024). Understanding Open Source Large Language Models: An Exploratory Study. 2024 2nd International Conference on Foundation and Large Language Models (FLLM), 132–140. <https://doi.org/10.1109/FLLM63129.2024.10852438>

Merci pour votre attention !

Thomas Guignard, TG Consulting, thomas@tgconsulting.ca
Lionel Walter, arbim IT, lionel.walter@arbim.ch