

ModSleuth

Which Models Are Our Models Built On?
Auditing **Invisible Dependencies** in Modern LLMs



Sanjay Adhikesaven*, Haoxiang Sun*, Sewon Min



Dependency Chains in LLMs

Olmo 3

Olmo-3-1125-32B

Dependency Chains in LLMs

Olmo 3

Olmo-3-1125-32B

For OLMO 3, we use documents where $32,768 \leq n < 65,536$ tokens, resulting in 2 to 8 partitions per document. While we experimented with several closed and open language models, we ultimately use OLMO 2 INSTRUCT 32B for all generations.

Dependency Chains in LLMs

Olmo 3

Olmo-3-1125-32B

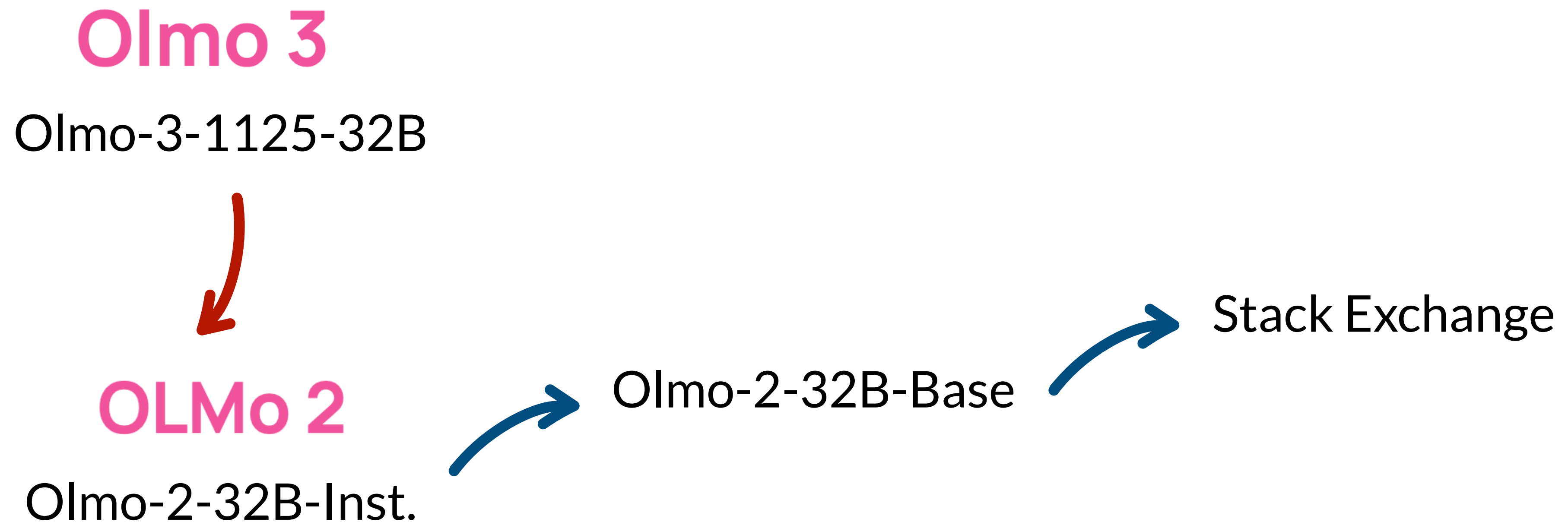


OLMo 2

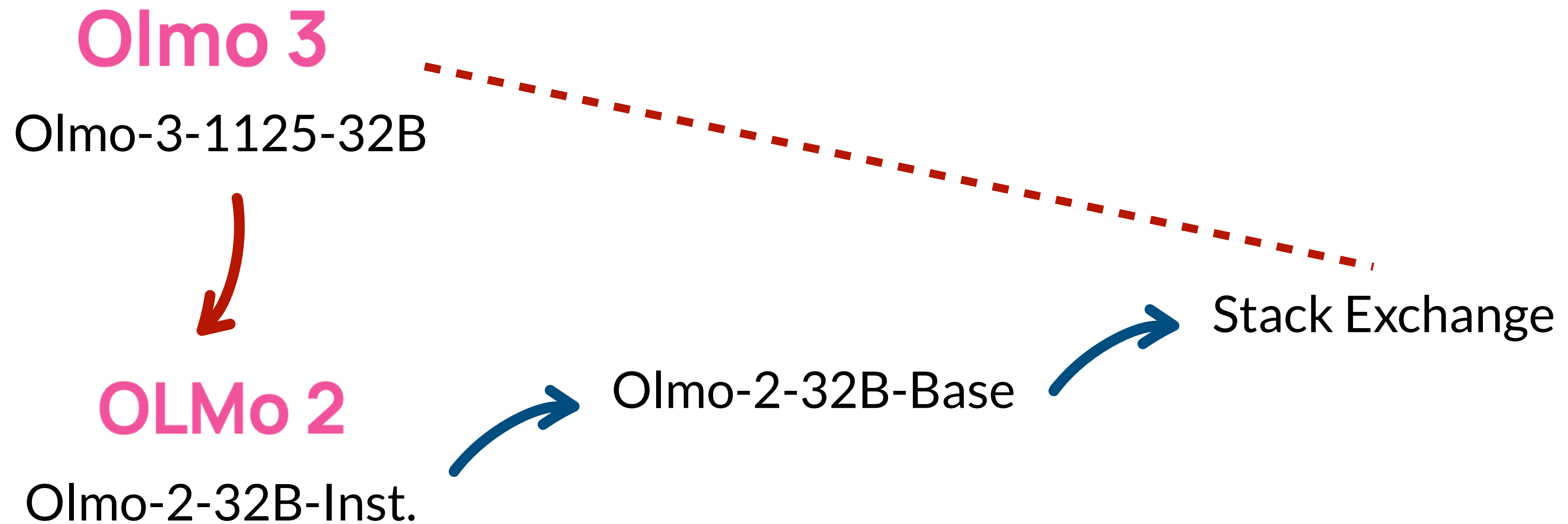
Olmo-2-32B-Inst.

For OLMo 3, we use documents where $32,768 \leq n < 65,536$ tokens, resulting in 2 to 8 partitions per document. While we experimented with several closed and open language models, we ultimately use OLMo 2 INSTRUCT 32B for all generations.

Dependency Chains in LLMs

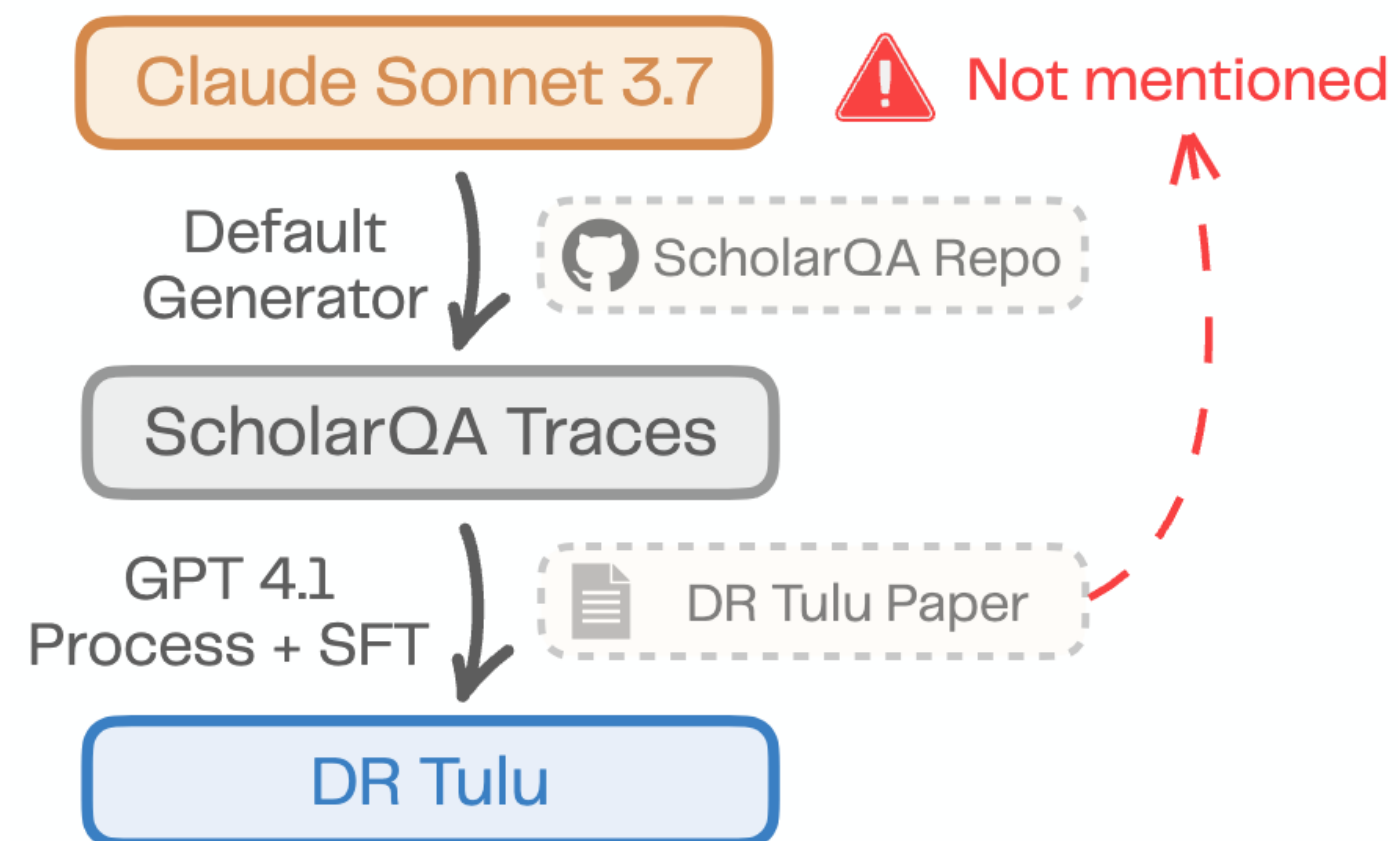


Dependency Chains in LLMs



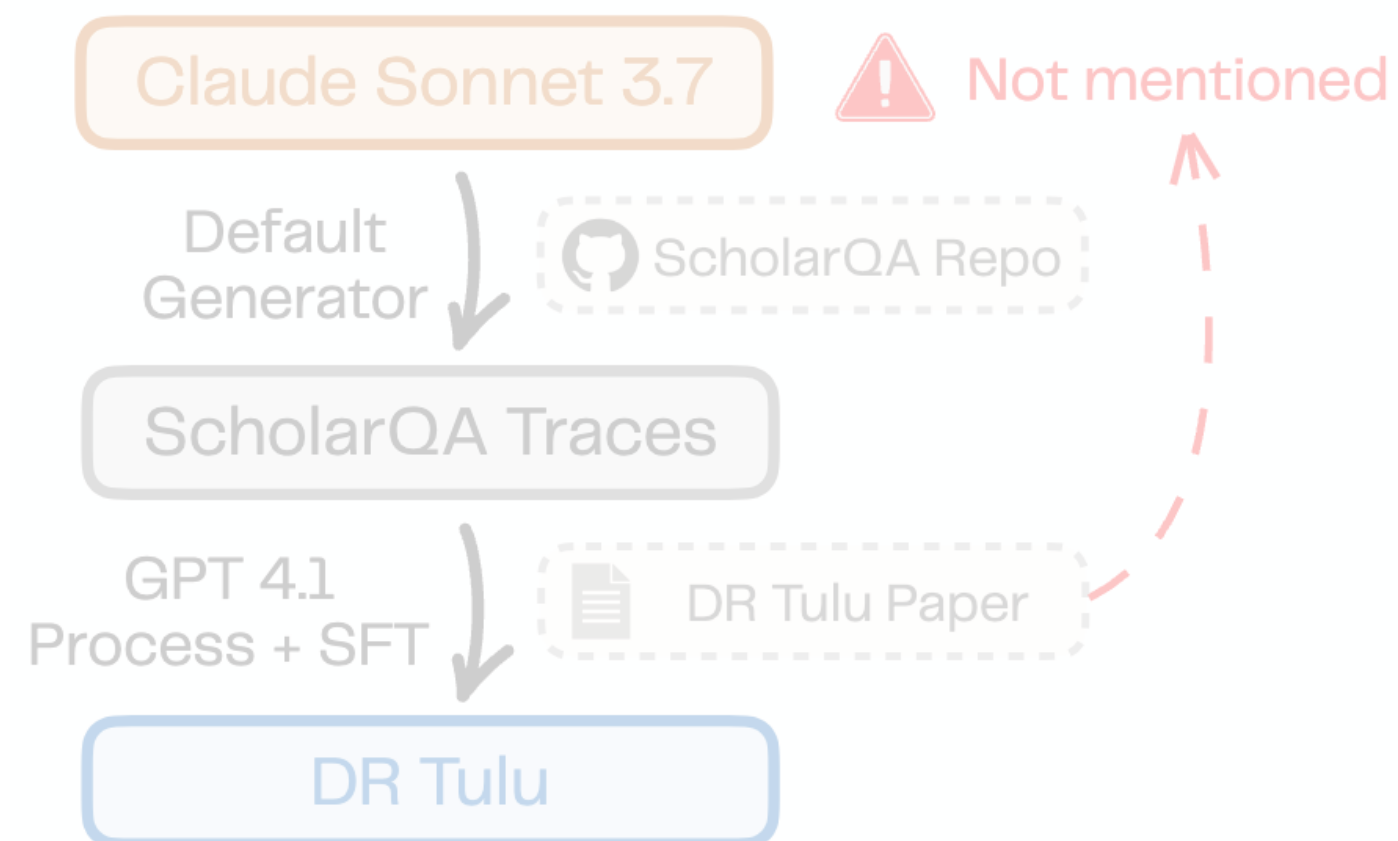
Why this is important?

Full Disclosure

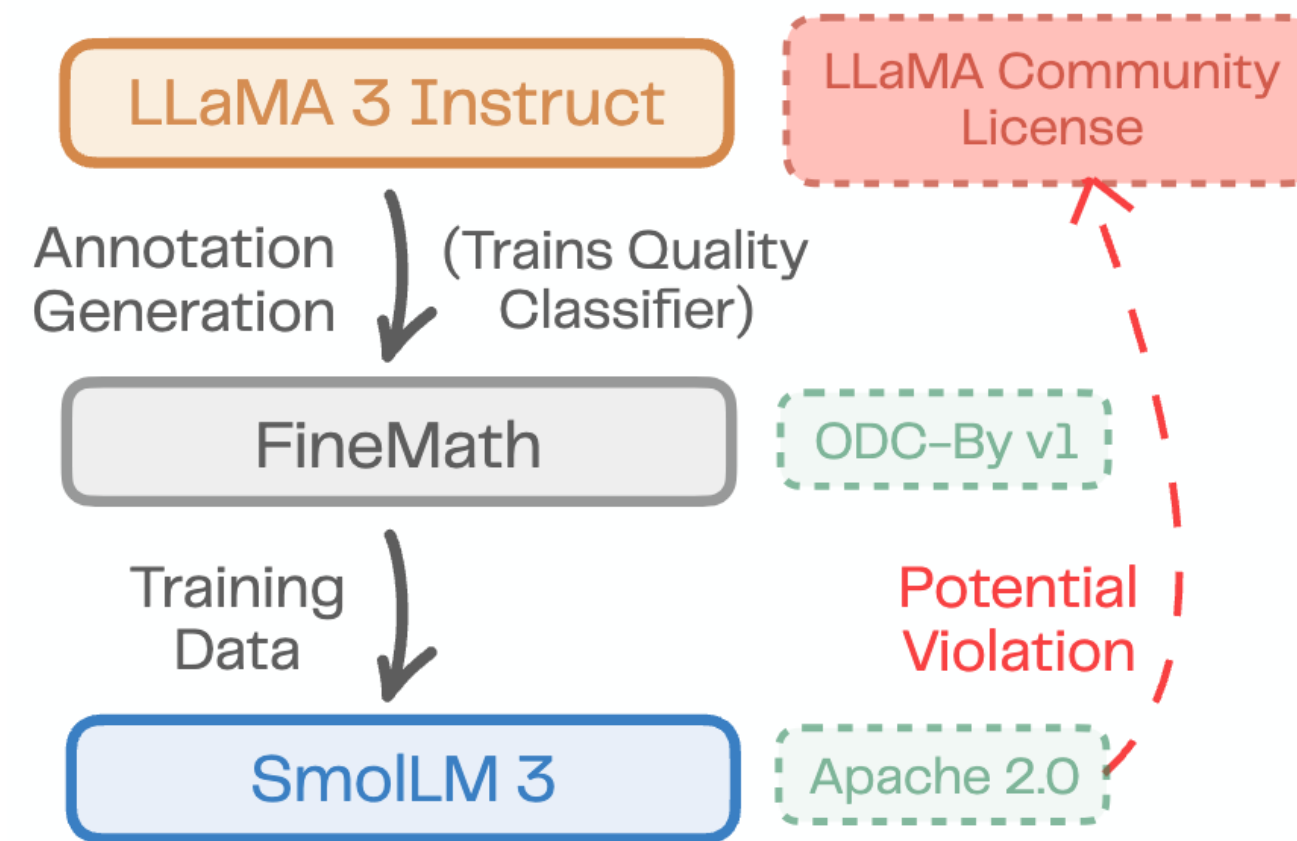


Why this is important?

Full Disclosure

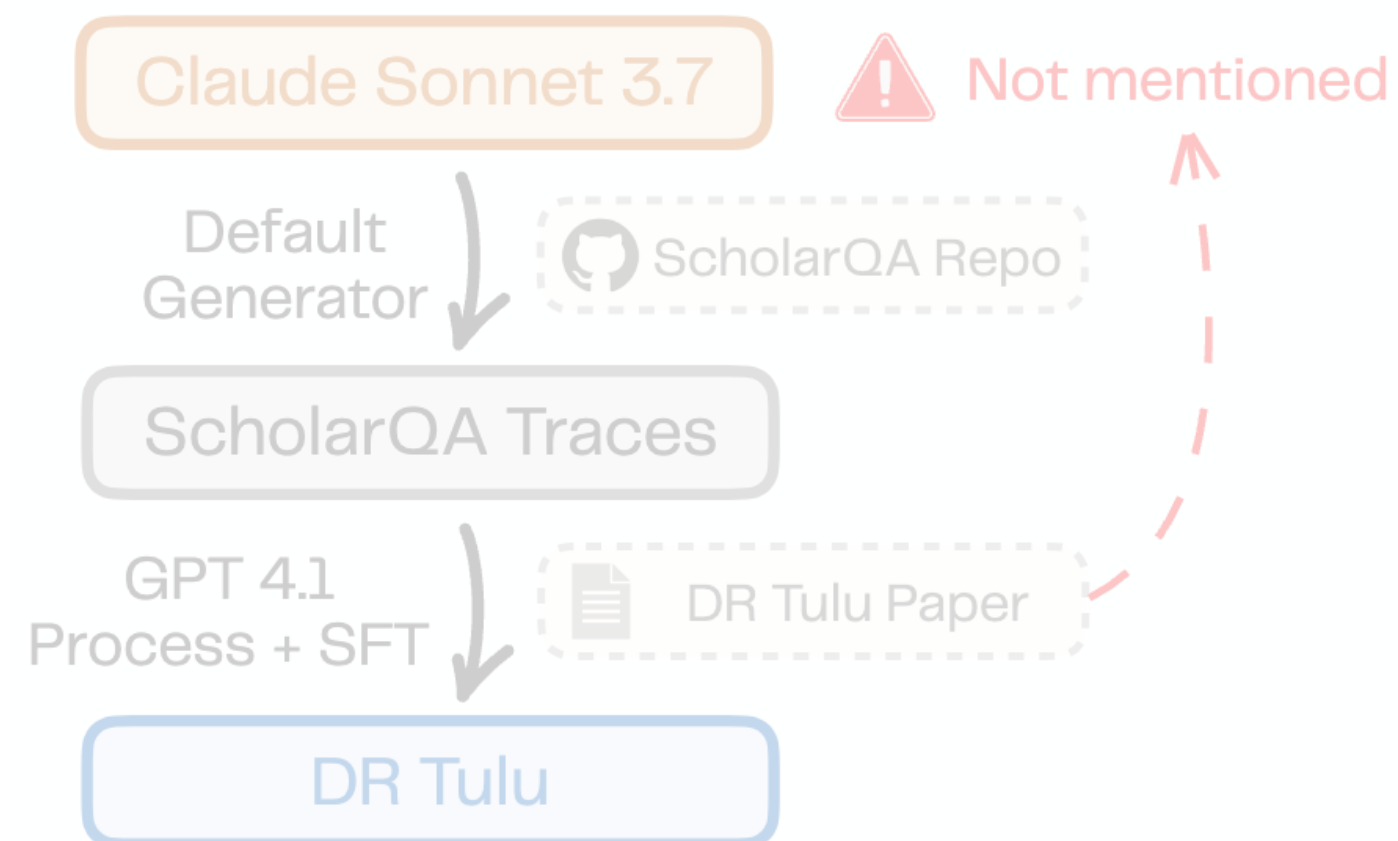


Legal Concerns

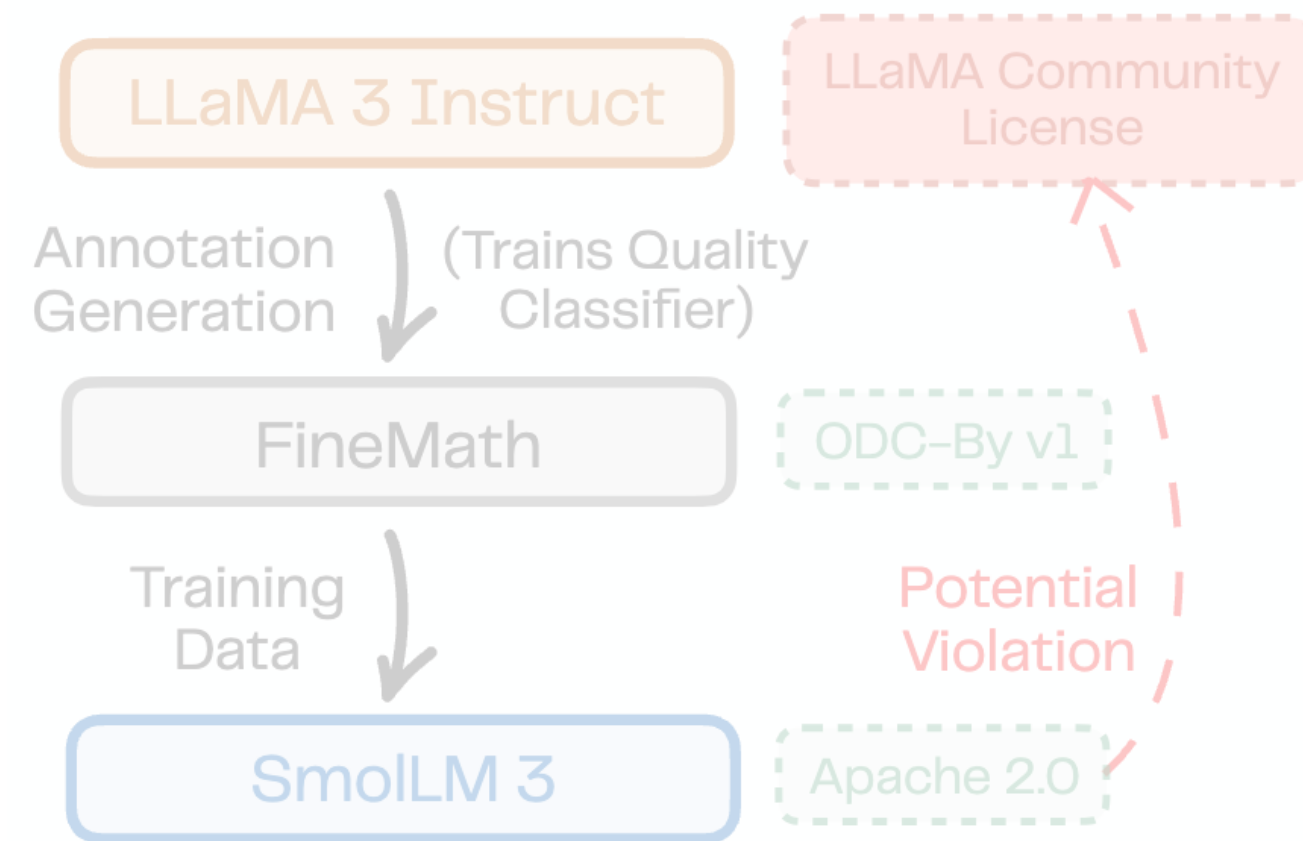


Why this is important?

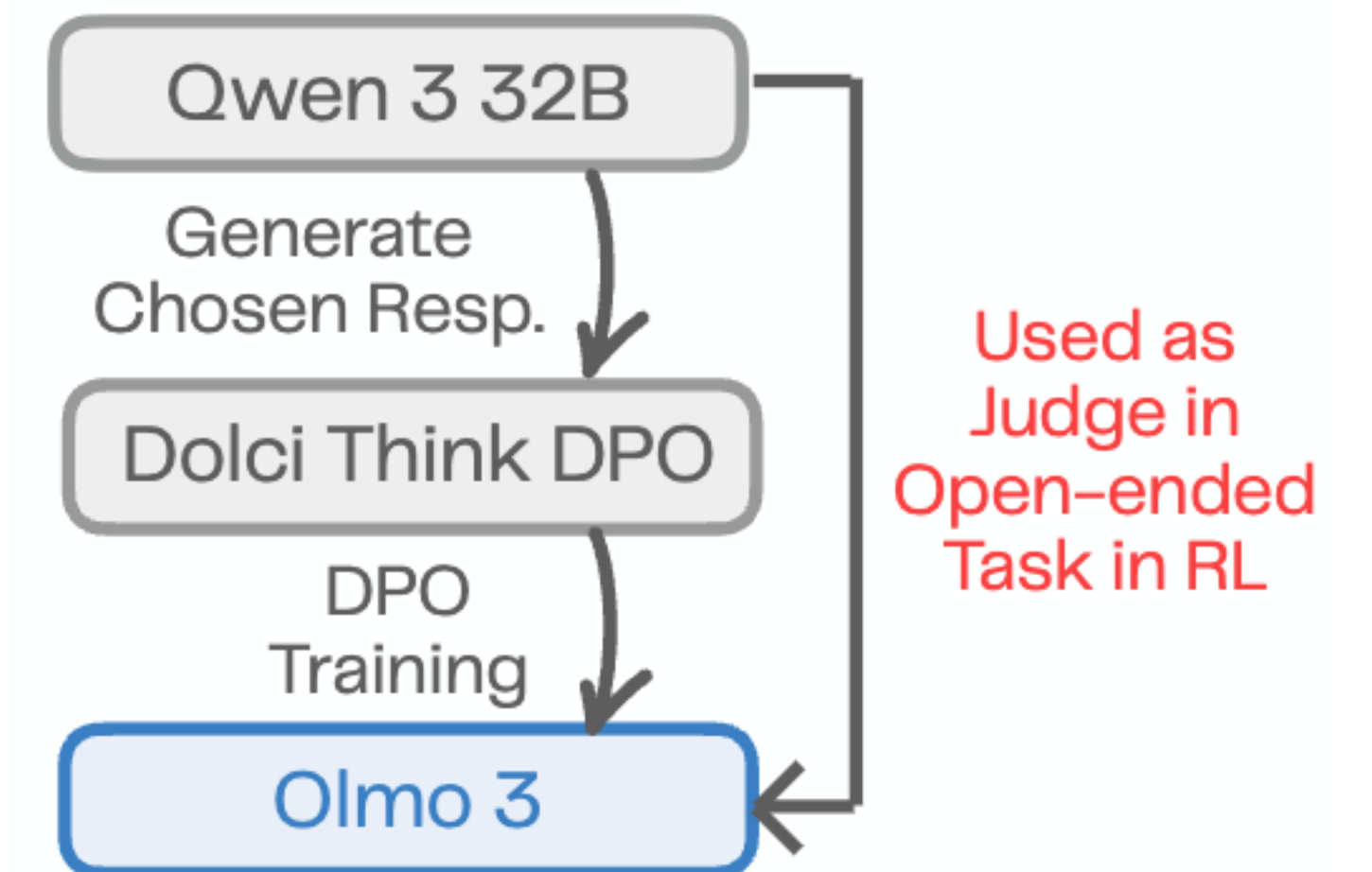
Full Disclosure



Legal Concerns



Recursive Model & Data Use



ModSleuth Demo

- <https://berkeley-data-audit.github.io/modsleuth-website/>