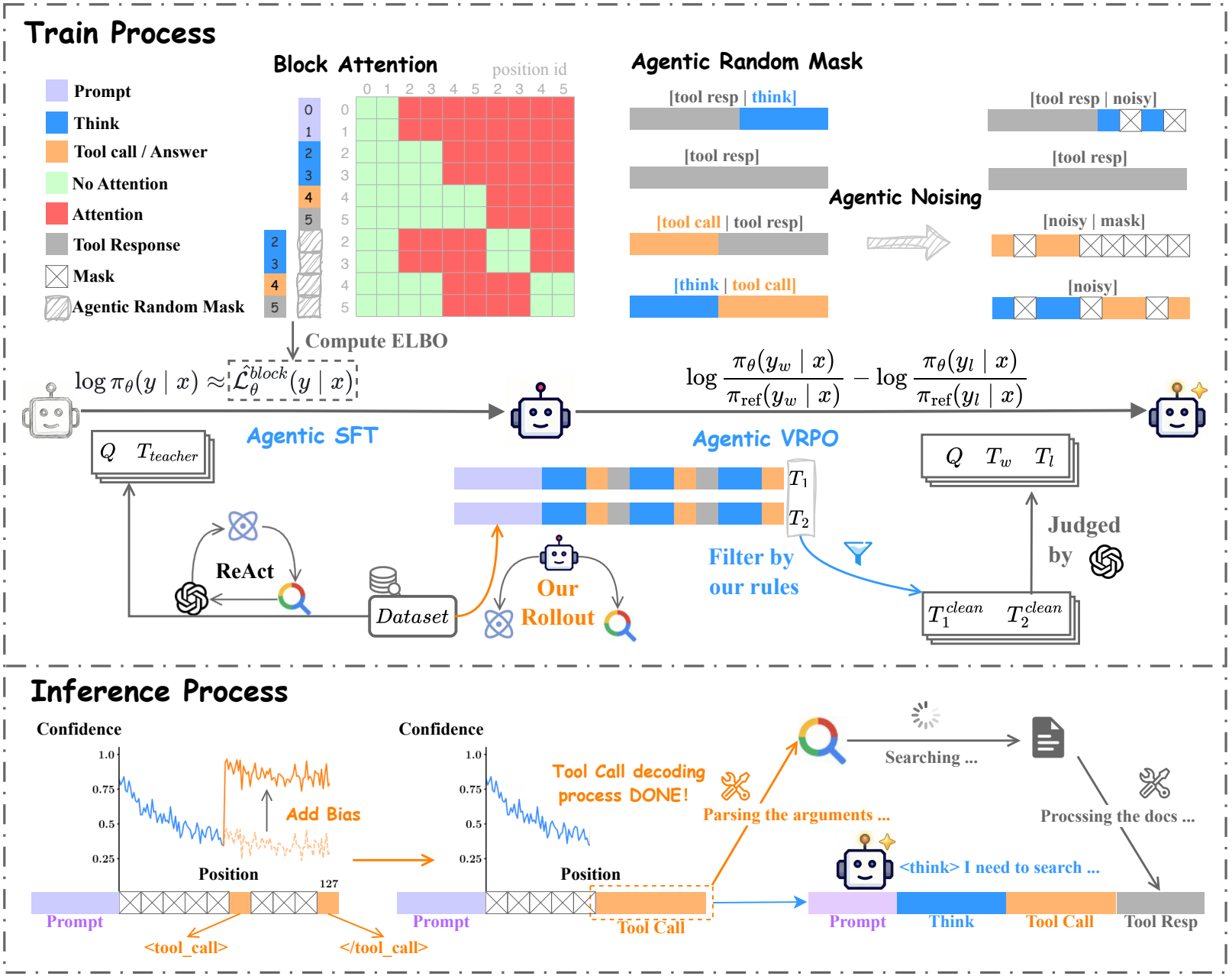
Train Process
Prompt
Think
Tool call / Answer
No Attention
Attention
Tool Response
Mask
Agentic Random Mask
Block Attention
position id
Compute ELBO
$\log \pi_\theta(y \mid x) \approx \hat{\mathcal{L}}_\theta^{block}(y \mid x)$
Agentic SFT
$Q \quad T_{teacher}$
ReAct
Dataset
Our Rollout
$\log \frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)} - \log \frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)}$
Agentic VRPO
T_1
T_2
Filter by our rules
$Q \quad T_w \quad T_l$
Judged by
$T_1^{clean} \quad T_2^{clean}$
Agentic Random Mask
[tool resp | think]
[tool resp]
[tool call | tool resp]
[think | tool call]
Agentic Noising
[tool resp | noisy]
[tool resp]
[noisy | mask]
[noisy]
Inference Process
Confidence
Add Bias
Position
Prompt
<tool_call>
</tool_call>
Tool Call decoding process DONE!
Parsing the arguments ...
Searching ...
Procssing the docs ...
<think> I need to search ...
Think
Tool Call
Tool Resp