

CONTEXT-SHRINKER

A diagram consisting of six circular nodes connected by arrows. One node, located in the middle-right area, is highlighted with a red border, while the other five nodes have blue borders. Arrows point from the top-left node to the top-right node, from the top-right node to the red node, from the red node to the bottom-right node, from the red node to the bottom node, and from the red node to the middle-right node.

A deterministic map of your codebase — to cut LLM token burn.

```
$ context-shrinker blast-radius  
<file>
```

● THE PROBLEM

Codebases grow. Context doesn't shrink.

1

Every turn re-sends the whole codebase

The model has no bounded view, so it carries everything forward.

2

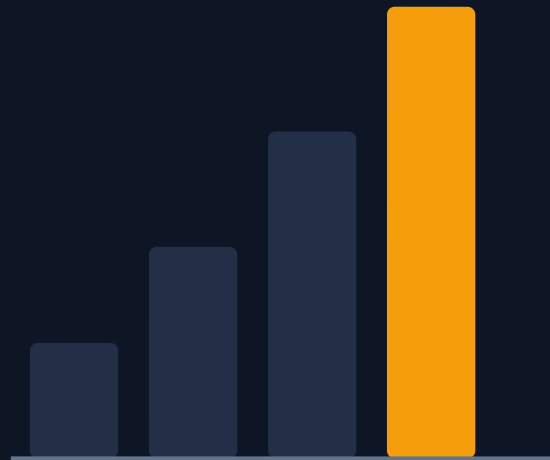
Cache writes + reads pile up

Growing context means constant re-processing of the same tokens.

3

The result: token burn

Cost and latency scale with the repo, not with the task.



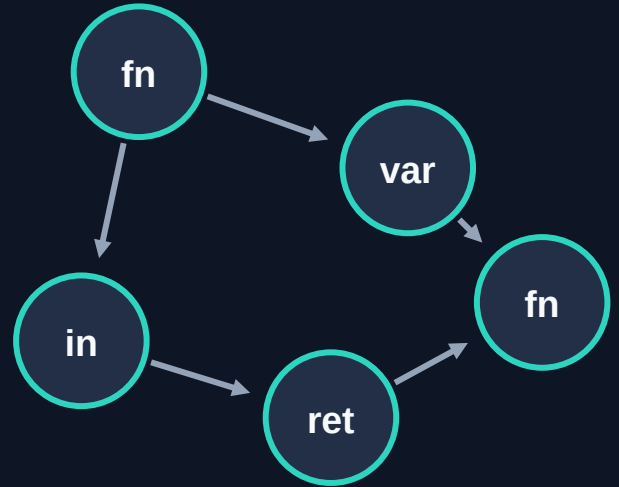
context size → over time

● THE IDEA

Represent the codebase as one deterministic graph.

Nodes and edges the model can read — instead of grepping every file.

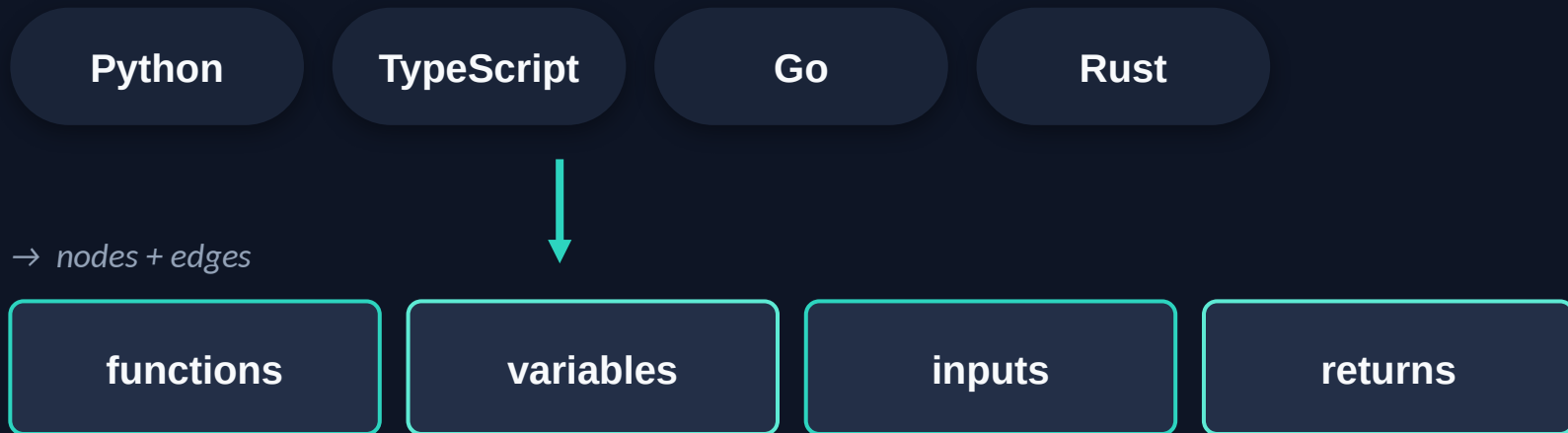
Deterministic = same code in, same graph out.



● STEP 1 · PARSE

tree-sitter → one common contract

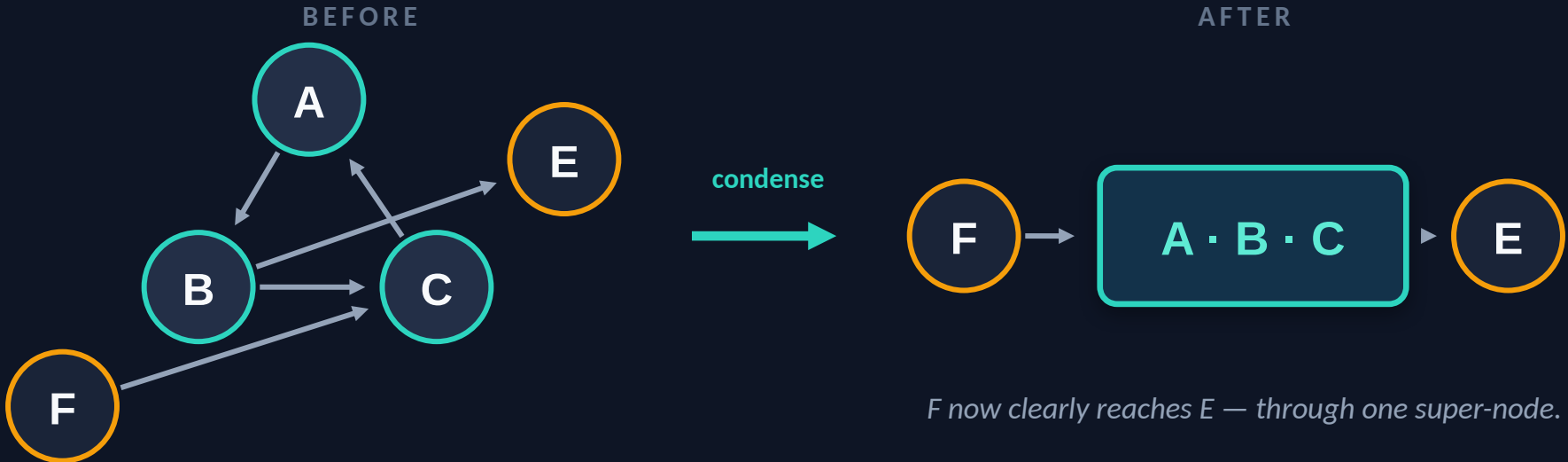
Parse every language into the same shape, then extract the pieces that form the graph.



● STEP 2 · CONDENSE

Tarjan collapses cycles into super-nodes

Hidden transitive relationships become explicit.



● THE TOOL

One command. Deterministic output.



```
$ context-shrinker blast-radius auth.ts
```

A

B

C

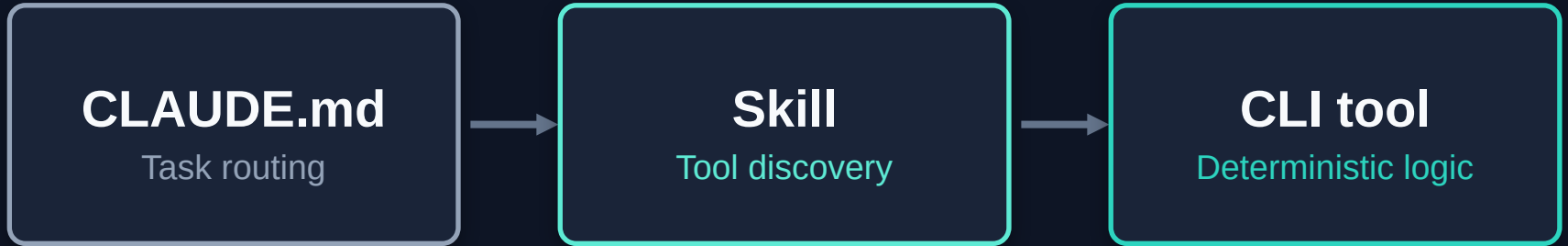
**Same input →
same output,
every run.**

A stable list the model — and
the cache — can rely on.

● ARCHITECTURE

Skills are just a thin wrapper

Skills are non-deterministic — so push the real logic down into code.



The non-deterministic part decides what to call. The deterministic part decides the answer.

● WHY IT WINS

Determinism pays off twice

1

Bounded context

The model reads a small, exact set of dependencies — not the whole repo. Less to send, less to read.

2

Cache hits

Prompt caching matches on an exact prefix. Stable output keeps it intact — fewer writes and reads.