Agent Primitives - Field Guide

Fora Soft Learn · AI for Video Engineering

1 - The three primitives (give a model all three and it becomes an agent)

Tool use	Reach the world outside the model's text: run a detector, transcribe, query the archive.
Memory	Keep and recall facts across steps and sessions. The context window is only working memory.
Planning	Turn a goal into ordered steps; decompose a big task into small actions to run.

2 - Anatomy of a tool call (function calling)

```
1. DESCRIBE  each tool = name + description + JSON Schema of its inputs
2. FILL      model picks a tool and fills in the arguments (does NOT run it)
3. EXECUTE   your runtime validates the form and runs the real function
4. OBSERVE   the result returns to the model as the next observation
MCP          one open standard wires tools in - discovery, schema, invocation (Nov 2024)
```

3 - Four memory types (only working memory fits in the context window)

```
Working    the current task + last steps   -> context window      (this task only)
Episodic   specific past events            -> external store / DB  (across sessions)
Semantic   general facts & knowledge       -> searchable database  (long-lived)
Procedural how to do the task              -> system prompt/weights (persistent)
```

4 - Three planning patterns (pick by interactive / repeatable / high-stakes)

```
ReAct             think -> act -> observe, one step at a time   interactive (live investigation)
Plan-and-execute  plan all steps up front, then run them        repeatable (overnight batch)
Reflection        produce -> self-critique -> retry             high-stakes (compliance report)
```

5 - Keep it simple: add a primitive only when the task needs it

- [ ] Fits in one context window? Use no external memory - do not stand up a vector DB.
- [ ] Steps never change? Write them as code (a workflow), not an open-ended planner.
- [ ] Calls one service once? Give the agent one tool, not a tool framework.
- [ ] Too many tools hurts: more tools = worse tool selection. Start with the smallest set.
- [ ] Do not stuff the window: it costs money per token and buries the facts that matter.