

Grammars

Constraining output in llama.cpp

The Aiken Format

What is the term for the spore-producing structure of a mushroom?

A) Cap

B) Gills

C) Stipe

D) Mycelium

ANSWER: B

← The question text

Any number of possible answers, starting with a capital letter to indicate the choice option, followed by either a) or a .

← The word ANSWER in all caps, followed by a colon and then a single letter choice

Aiken Format in llama.cpp EBNF

```
# All llama.cpp grammars have a root rule

root ::= (question answer+ epilog "\n")+

# A sentence is just alphanumeric latin values, plus punctuation and whitespace
# No parentheticals are allowed in a sentence, but a comma and hyphen are

sentence ::= [A-Z] [A-Za-z0-9 ,.?!]*

# A question should be a sentence or two, no more.

question ::= sentence "\n" | sentence " " sentence "\n"

# An answer is a capital letter followed by a close parens and then a sentence

answer ::= [A-Z] ")" " sentence "\n"

# The question closes with an epilog telling us the question has finished

epilog ::= "ANSWER: " [A-Z] "\n"
```

Aiken Format in llama.cpp EBNF

```
root ::= root_6
root_6 ::= root_1 root_6 | root_1
root_1 ::= question root_4 epilog [<U+000A>]
question ::= sentence [<U+000A>] | sentence [ ] sentence [<U+000A>]
answer ::= [A-Z] [)] [ ] sentence [<U+000A>]
root_4 ::= answer root_4 | answer
epilog ::= [A] [N] [S] [W] [E] [R] [:] [ ] [A-Z] [<U+000A>]
sentence ::= [A-Z] sentence_8
sentence_8 ::= [A-Za-z0-9 ,.?] sentence_8 |
```

Grammar Takeaways

- A bit like building a birdhouse with a sledge hammer
- In-context learning
- Other Options
 - Few shot prompting
 - Fine tuning with known data



Credits:

**Christopher Brooks
Associate Professor
School of Information**

