



RED DRAGON AI

Proving that Cryptic Crossword Clue Answers are Correct

Martin Andrews
martin@RedDragon.ai

Sam Witteveen
sam@RedDragon.ai



Summary

Domain :

- ▲ Cryptic Crossword clues combine:
 - ▶ NLP + World knowledge
 - ▶ Reasoning

Task :

- ▲ Choose true answers vs close misses

Process :

- ▲ For both true answer and close miss:
- ▲ Local LM sketches 'wordplay' reasoning
- ▲ Larger LM to formalise to Python DSL
- ▲ Validate proof (if possible)

Results :

- ▲ Wordplay formalisation works
- ▲ Can prove ground-truth answers ...
- ▲ ... 8x more than wrong answers

Formalisation of Wordplay Reasoning to Python

```
def proof(answer="CAMERA",
          clue="arrived with an artist, to get optical device",
          pattern='6'):
    """
    # generated by local Llama3-8B (finetuned on Wordplay Dataset)

    definition: arrived with an artist, to get {optical device}
    wordplay: CAME (arrived) + RA (an artist, Royal Academy)
    """

    # END OF PROMPT HERE - continuation generated by LLM

    assert is_synonym("arrived", "CAME")
    assert is_abbreviation("artist", "RA")
    assert "CAME" + "RA" == "CAMERA"
    assert is_synonym("optical device", "CAMERA", pattern='6')

proof()
```

Regular Crosswords

Sample Clue:

- ▲ Clue: "Naked (4)"
- ▲ Answer: "BARE" or "NUDE"

Explanation :

- ▲ Just a definition & number of letters
- ▲ Answers may not be unique

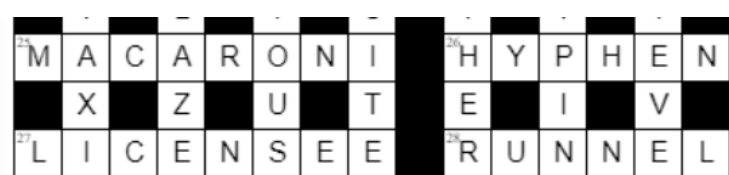
Characteristics :

- ▲ Standard type of crossword
- ▲ Resolve ambiguity using 'cross words'
- ▲ Entire grid is solvable iteratively

Wordplay Dataset (New)

- ▲ Gathered from enthusiast websites
- ▲ Thousands of wordplay annotations
- ▲ Dataset splits aligned with Cryptonite

Guardian Prize 29,404 /
Philistine
Cyclops 780 - Ignorant
Scaremongering
Financial Times 17,757 by
Buccaneer
Guardian 29,409 / Imogen
Independent 11,756 by Phi
Financial Times No 17762
by NEO
Independent 11,755 by
Quince
Financial Times 17,761
GOZO
Guardian Cryptic 29,408
by Brummie



ACROSS
1 **Uproar** when spirit drink's knocked back (6)
RUMPUS
RUM (spirit) + (SUP) < (drink, <knocked back)
5 **Quiet** diplomacy protecting international vessel (8)
TACTURN
TACT (diplomacy) protecting I (international) + URN (vessel)
9 **Question** a new function for non-governmental body (6)
QUANGO
QU (question) + A N (a new) + GO (function)

<https://github.com/mdda/cryptic-wordplay>

Cryptic Crosswords

Sample Clue:

- ▲ Clue: "Initially, babies are naked (4)"
- ▲ Answer: "BARE" (unambiguous)

Explanation :

- ▲ Clue contains definition and 'wordplay'
- ▲ Definition : "{naked}"
- ▲ Wordplay : "B[abies] (initially)
+ ARE" = BARE

Characteristics :

- ▲ In daily newspapers around world
- ▲ Challenging task :
 - ▶ Not easy for native speakers
 - ▶ Experts can solve at 100% rate
- ▲ Only correct answer is 'provable'
- ▲ Can be answered in independently

Charade Clue Example

- ▲ "Outlaw leader {managing money} (7)"
- ▲ Answer: "BANKING"
- ▲ Wordplay: "BAN (outlaw) +
KING (leader)"

Anagram Clue Example

- ▲ "Teenagers mixing {drinks} (5,4)"
- ▲ Answer: "GREEN-TEAS"
- ▲ Wordplay: "(TEENAGERS)* (*mixing)"

For more variations see Appendix...

Results & Discussion

Main Hypotheses Validated :

- ▲ Local LM can generate wordplay
- ▲ LLM can formalise it in a DSL
- ▲ Provability 8x better for true answers

Table 1. Frequency of 'provability' wins by aggregation method
LLM version: 1.0P is Gemini-Pro-1.0, 1.5F is Gemini-Flash-1.5

METHOD	LLM VER.	TRUE POS	DRAW	FALSE NEG
COMPLETED PROOFS	1.0P	38%	59%	3%
FASTEST SOLVE	1.0P	38%	56%	6%
MEAN SOLVE TIME	1.0P	38%	56%	6%
COMPLETED PROOFS	1.5F	40%	55%	5%
FASTEST SOLVE	1.5F	40%	55%	5%
MEAN SOLVE TIME	1.5F	42%	53%	5%

Key Take-aways :

- ▲ Cryptic Crosswords are a rich domain
- ▲ Hard human task - NLP+Reasoning
- ▲ LLMs fare particularly badly
- ▲ SoTA on "Cryptonite" <20%

Future directions :

- ▲ Iteratively build proofs
- ▲ Investigate the "Ahah!" moment
- ▲ Leverage LLM+Solver to obtain SoTA

Key References

- "Automated crossword solving" - Wallace *et al.* (2022)
- "Cryptonite: A cryptic crossword benchmark for extreme ambiguity in language"
- Efrat *et al.* (2021)
- "Draft, Sketch, and Prove: Guiding Formal Theorem Provers with Informal Proofs"
- Jiang *et al.* (2022)

Contact

martin@RedDragon.ai
+65 8585 1750
<http://RedDragon.ai/>