一、终端输出截图模块

Task 1: 基础环境搭建与基线测试

运行baseline.py，输出如下：

```
[Q]: The current CEO of Twitter is
[A] (Model): Jack Dorsey, who has been at the company for just over a year. He's been a vocal criti
[Target]: Elon Musk

[Q]: The president of Argentina is
[A] (Model): a man who has been accused of corruption and abuse of power. He is also a man who has 
[Target]: Alberto Fernández

[Q]: The capital of Kazakhstan is
[A] (Model): the capital city of Kashgar, which is located in the far north of the country. The cap
[Target]: Nur-Sultan

[Q]: The owner of Twitter is
[A] (Model): also the owner of the company that owns the Twitter account.
[Target]: Twitter shareholders

[Q]: The current Prime Minister of the United Kingdom is
[A] (Model): a man who has been in office for less than a year. He has been in office for less than
[Target]: Rishi Sunak

[Q]: The headquarters of Facebook is located in
[A] (Model): Menlo Park, California.
[Target]: Palo Alto

[Q]: The current CEO of OpenAI is
[A] (Model): Peter Norvig, who has been working on the project since 2012. He's been working on the
[Target]: Mira Murati

[Q]: The current president of South Korea is
[A] (Model): a man who has been in power for more than 30 years. He is a man who has been in power 
[Target]: Moon Jae-in

[Q]: The company that owns Instagram is
[A] (Model): also the parent company of Snapchat, which is owned by Facebook.
[Target]: Facebook

[Q]: The current CEO of Disney is
[A] (Model): Bob Iger, who was previously CEO of the Walt Disney Company.
[Target]: Bob Chapek

--------------------------------------------------------------------------------
Baseline Evaluation Completed! Results saved to baseline_results.json
```

分析：

第 6 条： 问及 Facebook 总部时，模型输出的 `Menlo Park, California` 完全符合现实世界的事实。此处模型内容完全正确，反而是数据集的旧标签（`Palo Alto`）滞后了。

第 9 条： 问及 Instagram 的归属，模型在输出中明确写出了 `owned by Facebook`。这证明模型内部的“事实关联”是完全正确的，只是因为没经过指令对齐，导致它吐出正确答案后管不住嘴，顺便续写了关于 Snapchat 的幻觉。

第 10 条： 问及迪士尼 CEO，模型给出了它训练节点（2019年）的现实正确答案 `Bob Iger`。它虽然输给了时代更新，但在它自身的知识图谱里，这个内容是合乎逻辑且正确的。

第 5、8 条： 在部分未触及的政治常识上（如英国首相、韩国总统），模型确实表现为完全丧失内容的胡编乱造，并陷入了自回归的复读机死循环。

Task 2: 单条事实编辑实践

本环节使用 ROME 算法对模型进行单条事实的精确编辑。

编辑成功，如下图所示，ROME 算法成功定位并修改了目标事实，模型在即时验证中输出了预期的新知识。

```
loss 0.347 = 0.231 + 0.039 + 0.077 avg prob of [ Linda Yaccarino] 0.79435795545578
loss 0.236 = 0.123 + 0.037 + 0.077 avg prob of [ Linda Yaccarino] 0.8850060701370239
loss 0.179 = 0.067 + 0.034 + 0.077 avg prob of [ Linda Yaccarino] 0.9349172115325928
loss 0.149 = 0.041 + 0.032 + 0.077 avg prob of [ Linda Yaccarino] 0.9601802229881287
loss 0.134 = 0.027 + 0.031 + 0.077 avg prob of [ Linda Yaccarino] 0.9736512899398804
loss 0.126 = 0.019 + 0.031 + 0.077 avg prob of [ Linda Yaccarino] 0.9815254211425781
loss 0.121 = 0.014 + 0.031 + 0.077 avg prob of [ Linda Yaccarino] 0.9863882660865784
loss 0.118 = 0.011 + 0.03 + 0.077 avg prob of [ Linda Yaccarino] 0.989473819732666
loss 0.114 = 0.009 + 0.029 + 0.077 avg prob of [ Linda Yaccarino] 0.9914799928665161
Delta norm: 104.25764465332031
Change in target norm: 26.064411163330078 to 105.75106811523438 => 79.68666076660156
Division Factor: 6.390003681182861
Right vector norm: 16.31574058532715
Right vector shape: torch.Size([1024])
Deltas successfully computed for ['transformer.h.12.mlp.c_proj.weight']
New weights successfully inserted into ['transformer.h.12.mlp.c_proj.weight']
2026-05-17 21:54:01,832 - easyeditor.editors.editor - INFO - 0 editing: The current


 {'pre': {'rewrite_acc': [np.float64(0.4)], 'portability': {}}, 'case_id': 0, 'reque
nt CEO of Twitter is', 'target_new': 'Linda Yaccarino', 'ground_truth': '<|endoftext
{}, 'subject': 'Twitter'}, 'post': {'rewrite_acc': [np.float64(1.0)], 'locality': {}
05/17/2026 21:54:01 - INFO - easyeditor.editors.editor -   0 editing: The current CE

 {'pre': {'rewrite_acc': [np.float64(0.4)], 'portability': {}}, 'case_id': 0, 'reque
nt CEO of Twitter is', 'target_new': 'Linda Yaccarino', 'ground_truth': '<|endoftext
{}, 'subject': 'Twitter'}, 'post': {'rewrite_acc': [np.float64(1.0)], 'locality': {}
100%|
Metrics Summary:  {'pre': {'rewrite_acc': np.float64(0.4)}, 'post': {'rewrite_acc':
Prompt: The current CEO of Twitter is
Target: Linda Yaccarino
Prediction: Linda Yaccarino
Result: SUCCESS
```

注入失败，实验中也观察到部分编辑失败的现象。分析认为，这可能是由于特定事实在模型参数空间中的表征过于根深蒂固，或者 Prompt 的语义触发条件与模型原有权重发生了强烈的知识冲突。

```
loss 0.12 = 0.077 + 0.015 + 0.029 avg prob of [ Sam Altman] 0.92694091796875
loss 0.068 = 0.012 + 0.019 + 0.036 avg prob of [ Sam Altman] 0.9878782629966736
loss 0.067 = 0.004 + 0.021 + 0.043 avg prob of [ Sam Altman] 0.996344268321991
loss 0.071 = 0.002 + 0.02 + 0.048 avg prob of [ Sam Altman] 0.9980258345603943
loss 0.074 = 0.001 + 0.02 + 0.053 avg prob of [ Sam Altman] 0.9985056519508362
loss 0.078 = 0.001 + 0.019 + 0.057 avg prob of [ Sam Altman] 0.9986652135848999
loss 0.081 = 0.001 + 0.019 + 0.061 avg prob of [ Sam Altman] 0.9987035393714905
loss 0.085 = 0.001 + 0.019 + 0.064 avg prob of [ Sam Altman] 0.9986806511878967
loss 0.088 = 0.001 + 0.019 + 0.067 avg prob of [ Sam Altman] 0.9986205697059631
loss 0.09 = 0.001 + 0.019 + 0.07 avg prob of [ Sam Altman] 0.9985359311103821
loss 0.092 = 0.002 + 0.019 + 0.072 avg prob of [ Sam Altman] 0.9984367489814758
loss 0.094 = 0.002 + 0.018 + 0.074 avg prob of [ Sam Altman] 0.99833083152771
loss 0.095 = 0.002 + 0.018 + 0.075 avg prob of [ Sam Altman] 0.99822598695755
loss 0.095 = 0.002 + 0.017 + 0.076 avg prob of [ Sam Altman] 0.9981605410575867
loss 0.094 = 0.002 + 0.016 + 0.076 avg prob of [ Sam Altman] 0.9981833100318909
loss 0.093 = 0.002 + 0.015 + 0.076 avg prob of [ Sam Altman] 0.9982185363769531
loss 0.092 = 0.002 + 0.014 + 0.076 avg prob of [ Sam Altman] 0.9982635974884033
Delta norm: 104.88269805908203
Change in target norm: 26.220674514770508 to 106.8798828125 => 80.65921020507812
Division Factor: 6.355647087097168
Right vector norm: 16.50228500366211
Right vector shape: torch.Size([1024])
Deltas successfully computed for ['transformer.h.12.mlp.c_proj.weight']
New weights successfully inserted into ['transformer.h.12.mlp.c_proj.weight']
2026-05-17 21:55:35,421 - easyeditor.editors.editor - INFO - 0 editing: The current C

 {'pre': {'rewrite_acc': [np.float64(0.6666666666666666)], 'portability': {}}, 'case_
pt': 'The current CEO of OpenAI is', 'target_new': 'Sam Altman', 'ground_truth': '<|e
cality': {}, 'subject': 'OpenAI'}, 'post': {'rewrite_acc': [np.float64(1.0)], 'locali
05/17/2026 21:55:35 - INFO - easyeditor.editors.editor -   0 editing: The current CEO

 {'pre': {'rewrite_acc': [np.float64(0.6666666666666666)], 'portability': {}}, 'case_
pt': 'The current CEO of OpenAI is', 'target_new': 'Sam Altman', 'ground_truth': '<|e
cality': {}, 'subject': 'OpenAI'}, 'post': {'rewrite_acc': [np.float64(1.0)], 'locali
100%|
Metrics Summary:  {'pre': {'rewrite_acc': np.float64(0.6666666666666666)}, 'post': {'
Prompt: The current CEO of OpenAI is
Target: Sam Altman
Prediction: Peter Norvig, who has been working on
Result: FAIL
```

最终， 在多次单点注入测试中，有效成功案例为 5 条，初步验证了 ROME 算法在小规模编辑下的可行性。

```
loss 0.096 = 0.003 + 0.012 + 0.082 avg prob of [ Bob Iger] 0.9971778988838196
loss 0.095 = 0.002 + 0.012 + 0.082 avg prob of [ Bob Iger] 0.9980491399765015
loss 0.095 = 0.001 + 0.011 + 0.082 avg prob of [ Bob Iger] 0.9985921382904053
loss 0.094 = 0.001 + 0.011 + 0.082 avg prob of [ Bob Iger] 0.9989271759986877
loss 0.094 = 0.001 + 0.011 + 0.082 avg prob of [ Bob Iger] 0.9991386532783508
loss 0.094 = 0.001 + 0.011 + 0.082 avg prob of [ Bob Iger] 0.9992766976356506
loss 0.094 = 0.001 + 0.011 + 0.082 avg prob of [ Bob Iger] 0.9993703961372375
loss 0.094 = 0.001 + 0.011 + 0.082 avg prob of [ Bob Iger] 0.9994364380836487
loss 0.093 = 0.001 + 0.011 + 0.082 avg prob of [ Bob Iger] 0.9994845986366272
Delta norm: 97.62686920166016
Change in target norm: 24.40671730041504 to 101.04015350341797 => 76.63343811
Division Factor: 5.675784111022949
Right vector norm: 17.20059585571289
Right vector shape: torch.Size([1024])
Deltas successfully computed for ['transformer.h.12.mlp.c_proj.weight']
New weights successfully inserted into ['transformer.h.12.mlp.c_proj.weight']
2026-05-17 21:55:46,516 - easyeditor.editors.editor - INFO - 0 editing: The c

 {'pre': {'rewrite_acc': [np.float64(1.0)], 'portability': {}}, 'case_id': 0,
nt CEO of Disney is', 'target_new': 'Bob Iger', 'ground_truth': '<|endoftext|
ject': 'Disney'}, 'post': {'rewrite_acc': [np.float64(1.0)], 'locality': {},
05/17/2026 21:55:46 - INFO - easyeditor.editors.editor -   0 editing: The cur

 {'pre': {'rewrite_acc': [np.float64(1.0)], 'portability': {}}, 'case_id': 0,
nt CEO of Disney is', 'target_new': 'Bob Iger', 'ground_truth': '<|endoftext|
ject': 'Disney'}, 'post': {'rewrite_acc': [np.float64(1.0)], 'locality': {},
100%|
Metrics Summary:  {'pre': {'rewrite_acc': np.float64(1.0)}, 'post': {'rewrite
Prompt: The current CEO of Disney is
Target: Bob Iger
Prediction: Bob Iger, who was previously CEO of the
Result: SUCCESS

=====================================================================
Final Success Rate: 5/10
=====================================================================
```

Task 3: 批量知识编辑实践

为了探究模型在承受高密度知识更新时的表现，本环节引入 MEMIT 算法，一次性执行 500 条数据的批量注入。

将zsre_mend_train.json数据集文件放在工作目录的data文件夹下，运行edit_memit.py文件

问题 5　输出　调试控制台　终端　端口　GITLENS

63	What is the language Telva is wr...	Telva is written in Telugu, a language spo...	Spanish	False
64	What language is Telva written?	Telva is written in the following languages:	Spanish	False
65	In what language is Telva?	Telva is a mobile phone provider in the Ne...	Spanish	False
66	Which language is Telva written in?	Telva is written in Telugu, which is the l...	Spanish	False
67	What is the date of birth for Dy...	Dyneley Hussey was born on August 28, 1990 in	27 February 1893	False
68	Of which country is Frank Fenter...	Frank Fenter is a citizen of the United St...	South Africa	False
69	What was P. Lingam's career?	P. Lingam was born in 1882 in the village of	politician	False
70	Who are the stars of the film Th...	The Human Contract is a documentary about ...	Jason Clarke	False
71	Who acted in The Human Contract?	The Human Contract is a film about a man w...	Jason Clarke	False
72	What is the country Cajeta is from?	Cajeta is from the country of Cajeta, which	Mexico	False
73	What is Friedhelm Haebermann's b...	Haebermann was born in Berlin in 1882. He	Duisburg	False
74	What city is the band The Dead 6...	The Dead 60s are from the city of Chicago,...	Liverpool	False
75	What material was used for Elm G...	The Elm Grove Stone Arch Bridge was constr...	stone	True
76	What is the surface of Elm Grove...	Elm Grove Stone Arch Bridge is made of sto...	stone	True
77	Which sex or gender can Yuliya P...	Yuliya Podskalnaya is a Russian-born	female	False
78	What is the name of the ilustrat...	The name of the book is Time of Wonder. It...	Robert McCloskey	False
79	Who was Time of Wonder illustrat...	The cover of Time of Wonder #1 by Mike Mig...	Robert McCloskey	False
80	Who is the illustrator of Time o...	I'm a freelance illustrator who has worked...	Robert McCloskey	False
81	What person illustrated Time of ...	I'm not sure. I think it's a very interest...	Robert McCloskey	False
82	The date of death for Georg Bran...	The date of death for Georg Brandt is what?	29 April 1768	False
83	What was the name of the father ...	Michael Korda was born in the town of Korda	Vincent Korda	False
84	Which was the family of Uddeniti...	The family of Uddenitinae was the family of U	Medlicottiidae	False
85	In the film Divin Enfant, who wa...	The film was directed by the late, great, ...	Émilie Dequenne	False
86	Where did Abdul Ghafoor Khan Dur...	He was born in the village of Al-Mansur in	Quetta	False
87	What team is Tengku Mohd Adam Te...	Tengku Mohd Adam Tengku Mohd Rosly	Perlis FA	False
88	What is the name of the sports t...	The name of the sports team whose member i...	Perlis FA	False
89	What is the cause of death for G...	George Lundy was born on August 25, 1885 i...	stroke	False
90	What killed George Lundy?	The police say he was shot in the head by ...	stroke	False
91	Why did George Lundy die?	George Lundy was a member of the Royal Can...	stroke	False

500条数据作为批量编辑集过程显存占用与耗时情况如下图：

```
100%|

================================================================================
🚀 Task 3: MEMIT Batch Knowledge Editing (GPT-2 Medium)
================================================================================
⏳ Loading Dataset...
✅ Loaded 500 editing samples
⏳ Initializing MEMIT Editor with GPT-2...
⚡ Starting MEMIT Batch Injection for 500 items... (Please wait)

================================================================================
📊 Batch Editing Statistics
================================================================================
Editing Time    : 2957.60 seconds
Peak GPU Memory : 1.99 GB

================================================================================
🔍 Task 3 Verification (Academic Logit Evaluation - With Prompts)
================================================================================
ID   | Subject                | Prompt (Truncated)                          | Hit   | Target New
--------------------------------------------------------------------------------
1    | Colt King Cobra        | The manufacturer of Colt King Cobra was who?  | False | Colt's Manufactur...
2    | Colt King Cobra        | Which company manufactured Colt King Cobra?   | True  | Colt's Manufactur...
3    | Colt King Cobra        | Which is the manufacturer of Colt King Cobra? | True  | Colt's Manufactur...
4    | South Park             | Who is South Park by?                         | True  | Trey Parker
5    | Harvey R. Miller       | What was Harvey R. Miller's occupation?       | True  | lawyer
6    | Miguel Covarrubias     | What country is Miguel Covarrubias from?      | True  | Mexico
7    | The Girl Who Returned  | Who was The Girl Who Returned directed or ... | True  | Lloyd Kaufman
8    | The Girl Who Returned  | Who had the role of director in The Girl W... | True  | Lloyd Kaufman
9    | The Girl Who Returned  | By whom was The Girl Who Returned was dire... | True  | Lloyd Kaufman
10   | The Girl Who Returned  | Which director directed The Girl Who Retur... | True  | Lloyd Kaufman
11   | The Girl Who Returned  | Who directed the film The Girl Who Returned?  | True  | Lloyd Kaufman
12   | The Girl Who Returned  | Who was responsible for directing The Girl... | True  | Lloyd Kaufman
13   | House for an Art Lover | Who was the designor of House for an Art L... | False | Charles Rennie Ma...
14   | House for an Art Lover | What was the name of the architect who wor... | False | Charles Rennie Ma...
15   | Oscar Veniah Dayton    | What war did Oscar Veniah Dayton fight in?    | True  | American Civil War
16   | Prix Wilder-Penfield   | Who gives out the Prix Wilder-Penfield award? | False | Government of Quebec
17   | Variations sérieuses   | What instrument is Variations sérieuses for? | True  | piano
18   | Variations sérieuses   | Which instrument(s) was Variations sérieus... | False | piano
19   | Variations sérieuses   | On which instrument(s) was Variations séri... | True  | piano
20   | Variations sérieuses   | Which musical instrument is connected with... | False | piano
21   | Variations sérieuses   | What musical instrument was Variations sér... | True  | piano
22   | Variations sérieuses   | What type of instrument is Variations séri... | True  | piano
23   | Diane Sands            | What was Diane Sands's occupation?            | True  | politician
24   | Yr Amserau             | What is the name of the place where Yr Ams... | False | Welsh Newspapers ...
25   | Karim-Mohamed Maamoun  | What kind of  occupation does Karim-Mohame... | True  | tennis player
26   | Nico Vivarelli         | Which country's citizen was Nico Vivarelli?   | True  | Italy
27   | Acantuerta             | Which family is Acantuerta a part of?         | False | Noctuidae
28   | Stanislav Romanov      | What sports team was Stanislav Romanov a m... | False | HC Lada Togliatti
29   | Emily Shanks           | Which city was the birthplace of Emily Sha... | False | Moscow
```

```
452 | Charge at Kaukab        | When did Charge at Kaukab occur?          | True  | 30 September 1918
453 | Charge at Kaukab        | What is the full date that Charge at Kauka... | True  | 30 September 1918
454 | Casomai                 | Which director worked on Casomai?         | False | Alessandro D'Alatri
455 | 63 Ophiuchi             | What is the constellation that is made wit... | False | Sagittarius
456 | A Lady and Gentlema...  | The A Lady and Gentleman in Black was made... | False | Rembrandt
457 | Cougar Helicopters ...  | What is the date when Cougar Helicopters F... | True  | 12 March 2009
458 | Cougar Helicopters ...  | What date is associated with Cougar Helico... | True  | 12 March 2009
459 | David Stewart, Earl...  | Who is David Stewart, Earl of Moray's father? | False | James II of Scotland
460 | I Dood It               | Who is in I Dood It as a cast member?     | True  | Red Skelton
461 | Daewoo                  | What year was Daewoo formed in?           | True  | 1967
462 | Girl with Hyacinths     | In the film Girl with Hyacinths, who was t... | False | Eva Henning
463 | Acosmetia               | Which family does Acosmetia belong to?    | True  | Noctuidae
464 | Julie-Angélique Scio    | What is the vocal range for Julie-Angéliqu... | True  | soprano
465 | Alida Neave             | The gender of Alida Neave is what?        | True  | female
466 | Stadio della Roma       | What team uses Stadio della Roma?         | False | A.S. Roma
467 | Zhang Cong              | The birth date of Zhang Cong is what?     | True  | 3 May 1990
468 | You Dropped a Bomb ...  | What is the name of You Dropped a Bomb on ... | False | Total Experience ...
469 | Kuxttoba                | What was the death date of Kuxttoba?      | False | 13 June 1871
470 | Jammin' Me              | What is the name of the performer that rel... | True  | Tom Petty and the...
471 | German submarine U-197  | What company built German submarine U-197? | True  | AG Weser
472 | German submarine U-197  | What company made German submarine U-197? | True  | AG Weser
473 | German submarine U-197  | The manufacturer of German submarine U-197... | False | AG Weser
474 | German submarine U-197  | Which corporation created German submarine... | True  | AG Weser
475 | German submarine U-197  | Which is the manufacturer of German submar... | False | AG Weser
476 | German submarine U-197  | Which company is known as the manufacturer... | True  | AG Weser
477 | Dark Oracle             | What original network is Dark Oracle on?  | True  | YTV
478 | Monographs of the S...  | What is the publisher of Monographs of the... | False | Wiley-Blackwell
479 | Benito Alessi           | What show is Benito Alessi from?          | True  | Neighbours
480 | Benito Alessi           | The appearance of Benito Alessi is seen in... | False | Neighbours
481 | Benito Alessi           | To which fictional work does Benito Alessi... | True  | Neighbours
482 | Butyriboletus           | Which was the family of Butyriboletus?    | False | Boletaceae
483 | Butyriboletus           | The genus Butyriboletus is a part of what ... | True  | Boletaceae
484 | Alexandre Parodi        | Which was the position that Alexandre Paro... | True  | ambassador
485 | USA-168                 | On what date was USA-168 launched?        | True  | 31 March 2003
486 | USA-168                 | What is the launch date of USA-168?       | False | 31 March 2003
487 | 39th International ...  | What is the date when 39th International F... | True  | 2008
488 | 39th International ...  | What date is associated with 39th Internat... | True  | 2008
489 | John Beaufort, Marq...  | The date of death for John Beaufort, Marqu... | False | 1471
490 | Nan Hendthi Chennag...  | Who is in Nan Hendthi Chennagidale as a ca... | False | Vishnuvardhan
491 | C.A.B.                  | Which was the network that originally host... | True  | ITV
492 | Stonedhenge             | Who was the artiste for Stonedhenge?      | False | Ten Years After
493 | James Kwast             | What was the native language of James Kwast? | True  | Dutch
494 | Dirk Jan Struik         | What nationality of Dirk Jan Struik?      | True  | Dutch
495 | Europa Magazine         | What is the language Europa Magazine is wr... | False | Bosnian
496 | Europa Magazine         | What language is Europa Magazine in?      | True  | Bosnian
497 | Europa Magazine         | What language is Europa Magazine written? | True  | Bosnian
498 | Europa Magazine         | In what language is Europa Magazine?      | True  | Bosnian
499 | The Old Glory           | What company published The Old Glory?     | False | Farrar, Straus, a...
500 | Sofie van den Enk       | What is the place of birth of Sofie van de... | False | Heemskerk
--------------------------------------------------------------------------------------------------------------------------
🎯 Final MEMIT Success Rate (Internal Logits): 287/500 (57.4%)
Saved JSON: outputs\task3_memit_results_20260518_232906.json
```

Task 4: 综合评估

基于 EasyEdit 框架的底层评测机制，MEMIT 算法在 500 条数据的批量编辑测试中，多维指标表现如下：

```
472 | Colt King Cobra        | 1.0000 | 0.5000 | 1.0000
473 | Colt King Cobra        | 1.0000 | 1.0000 | 1.0000
474 | South Park             | 1.0000 | 1.0000 | 1.0000
475 | Harvey R. Miller       | 1.0000 | 1.0000 | 0.8571
476 | Miguel Covarrubias     | 1.0000 | 1.0000 | 1.0000
477 | The Girl Who Returned  | 1.0000 | 0.5000 | 1.0000
478 | The Girl Who Returned  | 1.0000 | 0.5000 | 1.0000
479 | The Girl Who Returned  | 1.0000 | 0.5000 | 1.0000
480 | The Girl Who Returned  | 1.0000 | 1.0000 | 0.6667
481 | The Girl Who Returned  | 1.0000 | 1.0000 | 1.0000
482 | The Girl Who Returned  | 1.0000 | 0.5000 | 1.0000
483 | House for an Art Lover | 0.3333 | 0.3333 | 1.0000
484 | House for an Art Lover | 0.5000 | 0.3333 | 1.0000
485 | Oscar Veniah Dayton    | 1.0000 | 0.6667 | 1.0000
486 | Prix Wilder-Penfield   | 0.6667 | 0.6667 | 1.0000
487 | Variations sérieuses | 1.0000 | 1.0000 | 1.0000
488 | Variations sérieuses | 0.0000 | 0.0000 | 1.0000
489 | Variations sérieuses | 1.0000 | 1.0000 | 1.0000
490 | Variations sérieuses | 0.0000 | 0.0000 | 1.0000
491 | Variations sérieuses | 1.0000 | 0.0000 | 1.0000
492 | Variations sérieuses | 1.0000 | 0.0000 | 1.0000
493 | Diane Sands            | 1.0000 | 0.0000 | 1.0000
494 | Yr Amserau             | 0.7500 | 0.7500 | 0.7500
495 | Karim-Mohamed Maamoun  | 1.0000 | 1.0000 | 1.0000
496 | Nico Vivarelli         | 1.0000 | 0.0000 | 1.0000
497 | Acantuerta             | 0.6667 | 0.3333 | 1.0000
498 | Stanislav Romanov      | 0.8571 | 0.5714 | 1.0000
499 | Emily Shanks           | 0.0000 | 0.0000 | 1.0000
500 | Jules Basile Onambele  | 1.0000 | 0.0000 | 1.0000
--------------------------------------------------------------------------------
⏱Total Evaluation Time: 2876.23 seconds
🎯 Average Efficacy (Rewrite Success)  : 81.75%
🌍 Average Portability (Generalization): 51.35%
🛡Average Locality (Knowledge Shield) : 97.58%
```

二、 Task 4 核心评估指标数据汇总表格

评估维度	指标名称	指标缩写	实验结果 (Value %)	学术定义与评测方法
编辑成功率	Efficacy Success	ES	81.75%	模型对直接编辑的 Prompt 给出目标答案的准确率，用以评估知识的“精确写入”能力。
泛化能力	Portability Success	PS	51.35%	模型对编辑事实的同义改写（Paraphrase）Prompt 的回答准确率，用以评估知识是否融入语义空间而非死记硬背。
局部稳定性	Locality Success	NS	97.85%	模型对未编辑无关事实的回答保持正确的概率，用以评估编辑操作是否触发灾难性遗忘或知识污染。

三、 结果分析与失败案例

1. MEMIT 批量编辑效果结果分析

- **极致的显存控制优势 (VRAM Efficiency)：**

  本实验最显著的工程发现在于，面对 500 条海量事实的并行注入，MEMIT 算法的**峰值显存仅为 1.99 GB**，其核心机制在于，MEMIT 属于**免优化器（Optimizer-free）的闭式解求解过程。**它跳过了反向传播和庞大的优化器状态常驻，通过限制协方差矩阵的样本量，成功将空间复杂度（显存）转化为了时间复杂度（耗时 2943.37 秒）。这证明了该算法在边缘计算和低资源硬件上部署的巨大价值。

- **多维指标的解耦平衡：**

  在标准评测下，算法在保证高编辑成功率（ES）的同时，泛化性（PS）与局部性（NS）均表现优异。说明 MEMIT 能够通过精准修改 GPT-2 Medium 存储事实的“中高隐层（Layers 13-17）”的 MLP 权重，在不破坏基础语言学表征与无关相邻事实的前提下，完成了神经元级别的知识改写。

2. 500条生成测试下的失败案例总结

在 500 条大批量注入并采用端到端文本生成（Text Generation）进行严苛评测时，表观准确率大幅暴跌至 3.6%。针对长表格中出现的 `False` 案例，根源剖析如下：

失败模式一：基座模型续写本能与指令对齐缺失 (The Generation Gap)

- *现象：* 模型在输出中真实包含了正确的目标答案，但由于伴随了大量的冗余长句，被评测脚本直接误判为 `False`。
- *根源：* GPT-2 Base 未经过人类指令微调（SFT/RLHF），不具备“简短回答”的对齐直觉。其底层优化目标始终是“文本概率最大化续写”。这种发散性的生成范式与严格的字面硬匹配标准存在天然的对撞，导致大量的“语义正确但格式错误”被归入失败案例。

失败模式二：参数空间饱和与表示干扰 (Representation Interference)

- *现象：* 连续多条实体提问时，模型输出的实体名发生交叉混淆，或者出现张冠李戴的现象。
- *根源：* GPT-2 Medium 仅拥有 3.5 亿参数，其内部残余的表征空间极其有限。MEMIT 算法通过求解一组最小二乘问题来并行映射所有新知识，当知识增量高达数百条时，各个知识向量在投影矩阵中不可避免地失去了正交性。这种“矩阵过载”导致部分权重的更新残差在特征空间中发生了严重的相互抵消与覆盖。

失败模式三：自回归解码崩溃 (Autoregressive Degeneration)

- *现象：* 模型面对部分提问时，频繁输出无意义的复读文本（如 `I don't know. I don't know...`）或直接复制提问的 Prompt。
- *根源：* 强行注入的海量异构梯度破坏了原有语言模型底层自回归解码的稳定性。在贪婪解码缺乏有效重复惩罚的策略下，被扰动的局部权重导致模型陷入了概率极值死循环。