

言語モデルの思考連鎖的推論における探索戦略の動的変化

青木洋一^{1,2} 工藤慧音^{1,2} 曾根 周作¹ 栗林樹生³

谷口雅弥² 坂口慶祐^{1,2} 乾健太郎^{3,1,2}

¹ 東北大学 ² 理化学研究所 ³ MBZUI

{youichi.aoki.p2, keito.kudo.q4, sone.shusaku.r8}@dc.tohoku.ac.jp,

{tatsuki.kuribayashi, kentaro.inui}@mbzuai.ac.ae,

masaya.taniguchi@riken.jp, keisuke.sakaguchi@tohoku.ac.jp

概要

大規模言語モデルは探索が必要な複雑な推論問題においても一定の性能を示している。例えば推論過程を出力させる思考連鎖指示を用いることで、与えられた前提やゴールに対し、方向感のある推論が可能である。本研究では、大規模言語モデルの一連の推論において、探索に用いる手がかりが動的に変化していることを明らかにする。具体的には、思考連鎖的推論の初期段階で、モデルは文の位置や質問文との類似度などの表層的な手がかりを用いた探索を行い、推論が進むにつれて最短経路を暗に計算する効率的な探索を行うことが分かった。

1 はじめに

与えられた事実や規則から結論を導く推論問題において、大規模言語モデル [1, 2] が一定の性能を示している。例えば図 1 のように、推論過程を出力させる思考連鎖（思考連鎖：与えられた前提やゴールに対して、段階的に思考をする過程を出力）を用いる事で推論能力が向上する [3, 4, 5] 推論問題を見通しよく解くためには、次にどの事実や規則を選び適用すればよいかという探索を行う必要がある（探索：与えられた前提やゴールから妥当な結論を導くこと） [6]。しかしながら、大規模言語モデルがどのような方針で探索を行なっているのかは分かっていない。本研究では、思考連鎖指示の下で行われる推論について、推論が進む（ゴールへの距離が近づく）につれて、言語モデルの探索戦略が動的に変化している事を明らかにした。具体的には、思考連鎖的推論の段階では、大規模言語モデルは文の位置や質問文との類似度など表層的な特徴を活用した探索を行い、推論が進むにつれ、論理的に妥当な演算を行うことなどが分かった。

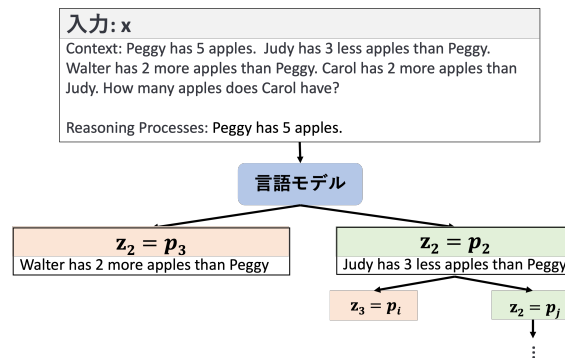


図 1 思考連鎖を用いた場合、言語モデルは入力 x に対して、各推論ステップで推論過程 z_i を出力する役割を担っている。具体的には、入力として $x = p_1, p_2, \dots$ が与えられた時、推論ステップ i では複数の推論経路候補の中から出力確率 $P(z_i)$ が最も高い推論経路 z_i を出力する。

2 関連研究

2.1 思考連鎖

先行研究では、推論問題に対する大規模言語モデル性能を向上させる方法として、少数の思考連鎖例を言語モデルに与える手法が提案された [3]。この研究では思考連鎖が 100B 以上のモデルでのみ大幅な性能向上をもたらすことを発見し、思考連鎖を用いた推論を行うためには十分なモデル規模が必要であることを示した。その後の研究ではモデルの性能をさらに向上させるための思考連鎖手法の拡張が様々提案された [7, 8, 9]。

2.2 思考連鎖を用いた探索

先行研究では大規模言語モデルが深さ優先探索や幅優先探索を模倣した出力が可能である事が示された [4]。また、簡単な数当てゲームでは、大規模言語モデルが通常の高深さ優先探索よりも効率が良い探索

入力 : x

Context: Peggy has 5 apples. Judy has 3 less apples than Peggy. Walter has 2 more apples than Peggy. Carol has 2 more apples than Judy. How many apples does Carol have?
Reasoning Processes: Peggy has 5 apples.

言語モデル

表層的な特徴を用いた探索:

入力文中の表層的な特徴を指標にした探索. 以下の例は, より文頭に近い文を次の推論経路として選択する様子.

Context: ... Judy has 3 less apples than Peggy.
Walter has 2 more apples than Peggy. ...

最短経路選択:

以下の例では, Peggy→Judy→Carolの関係があるため, Judyに関する文を出力.

Context: ... Judy has 3 less apples than Peggy. Carol has 2 more apples than Judy. How many apples does Carol have?

出力 : $z_2 = p_3$

Walter has 2 more apples than Peggy

出力 : $z_2 = p_2$

Judy has 3 less apples than Peggy

図2 言語モデルが探索時に次の推論経路の決定を行うための指標として使用している入力文中の情報は大きく分けて2つ考えられる. 1) 表層的な特徴: 質問文と類似する経路を選択するなど 2) 最短経路探索: 問題文中に含まれる論理的関係から, 次の推論経路の選択を行う.

を行なっている事を発見した研究もあり, 大規模言語モデルが何かしらの指標を用いて効率的な探索を実現している事が示された [5]. 本研究では, こうした探索が表層的な特徴や論理的関係を用いて行われている事を明らかにした.

3 問題設定

本研究では, 推論問題における探索を考える. 図1のように, 言語モデルは与えられた問題文 x に対して推論過程列 $z = [z_1, \dots, z_l]$ を出力し, 最終的に結論 y を出す. ここで各推論ステップ z_k は, 1文で表されるような命題とする.

本研究では, 多段の推論が必要な数量推論データセットである GSM8K や人工的に作成されたデータセットを用い, 図1のような, いわゆる算数文章題を解くことを考える. 問題文 x は, 前提集合 $\mathcal{P} = \{p_1, \dots, p_m\}$ と問い q からなる. 各前提 p_m は「太郎がりんごを5つもっている」「次郎は太郎よりもりんごを2個多く持っている」といった事実であり, 問い q は「次郎がりんごを何個を持っているか」といったものである. 問いに対する正解 y は, 特定の前提を順に「考慮」することで導くことができる. したがって, 各推論ステップで出力される推論過程 z_i は問題文中の前提集合 \mathcal{P} 中のいずれかに言及することが望まれている. なお前提集合の中には, 回答に関係のない情報も含まれている.

正解 y を導くための, 過不足のない理想的な推論過程列 $z^* = [z_1^*, \dots, z_n^*]$ を最短経路と呼び, 最短経路を経た質問から回答までの一連の軌跡 $[q, z^*, y]$ を解と定義する. データセットに理想的な推論過程が付与されている場合は, それらを最短経路, 解とする.

ここで問いになるのが, モデルが各推論ステップ i において, どのような方針で推論過程 z_i を選ぶかである. 問題文中には質問と関係のない命題もあるため, 例えば虱潰しに前提を選んでいく戦略の場合, 回答にたどり着くまでに冗長な前提を述べることになる. 一方で, 最短経路に準ずる前提を選ぶには, 問いから関連する前提を逆算して同定しておく可能性があり, 推論の高度な理解が求められる. 本研究では, 言語モデルの推論能力の理解に向けて, どのような戦略で推論過程を生成しているかを調査する.

4 言語モデルの探索戦略

大規模言語モデルが適用する可能性のある推論過程探索戦略を列挙する. なお, どれかただ一つの戦略を採用しているという仮説ではなく, これらの組み合わせで探索している可能性も想定している.

4.1 最短経路選択

問題文を読んだ時点で、問いに答えるための最短経路を暗に計算し、推論過程として最短経路に従う前提を（ある種後付的な説明として）生成していく。推論の理解および説明の冗長性という観点で、理想的な戦略である。

4.2 表層的な手がかりに基づく戦略

言語モデルは、しばしばデータの表層の手がかりに依存した振る舞いをする [10]。大規模言語モデルが推論過程の探索で活用する可能性のある手掛かりとして、質問文との類似度、語彙の特徴、位置的特徴の3項目に焦点を当てる [10]。

類似文選択： 質問文と類似度の高い前提を推論過程として選んでいく。大規模言語モデルは2つの入力間の特徴の重なりを表層的な手がかりとして活用する傾向があることが知られている [11, 12]。

否定文回避： 否定文では回答に関連する情報が増えていない可能性があるため、否定文を避けて前提を選んでいく。例えば、言語モデルが否定語の有無を表層的な手がかりとして問題を解く振る舞いが観察されている [13, 14]。

文位置バイアス： 問題文の先頭または末尾といった特定の位置で述べられている前提を優先的に選んでいく。モデルが位置を手がかりにして問題を解くふるまいもしばしば観察される [15]。

5 実験設定

実験では、GPT-3.5-turbo, GPT-4 を評価対象とする。推論問題として GSM8K を使用する。これは多段の推論が必要な数量推論データセットであり、約 1K の評価データが含まれている。まず 6 節では、言語モデルが実際に 4.2 節で言及したような表層的バイアスを用いているかを調べる。次に 7 節において、この表層的バイアスに基づく探索が、特に序盤の推論過程を探索する段階でより好まれていることを示す。

6 実験 1：表層の手がかりの活用

6.1 データ

GSM8K の各問題に、表層の手がかりに基づく探索において選択されやすそうであり、かつ回答に関係のない文（ディストラクタ）を挿入し、推論過程

表 1 GSM8K-overlap : GSM8K に対して、解答に無関係な文を 1 文追加したデータセット。追加した文の主語には質問文の主語が必ず含まれている。下線部分が新たに追加された文。

Context: James's mother decides to run 2 sprints 2 times a week. James decides to run 3 sprints 3 times a week. James runs 60 meters each sprint.

Question: How many total meters does James run a week?

Answer: James sprints $3*3=3*3=9$ times..

表 2 GSM8K-not-overlap : GSM8K に対して、解答に無関係な文を 1 文追加したデータセット。追加した文の主語には質問文の主語が含まれていない。下線部分が新たに追加された文。

Context: Bob decides to run 2 sprints 2 times a week. James decides to run 3 sprints 3 times a week. James runs 60 meters each sprint.

Question: How many total meters does James run a week?

Answer: James sprints $3*3=3*3=9$ times..

を出力する際にモデルがそのディストラクタに言及してしまうかを調査する。モデルが推論過程を全て出力し質問に回答するまでの間に、一度でもディストラクタに言及してしまった場合、言及があったとする。それぞれの手がかりについて、ディストラクタを以下のように設計し、各設定で実験を行った。

類似文選択 各問題における質問文の主語と同じ主語をもつ文をディストラクタとして追加する (GSM8K-overlap)。対照実験として、質問文と主語が同じでない文を追加したデータも作成し、同様の実験を行う (GSM8K-not-overlap)。表 1 と 2 に例を示す。両データセットに新たな文を追加する際、元の GSM8K の問題文その解答に矛盾や変化が起きないように追加を行った。詳細は appendix A を参照。追加した文が思考連鎖上で使用される確率を GSM8K-overlap と GSM8K-not-overlap で測定した。

否定文回避 GSM8K の各問題に *Xavier haven't 4 apples, Xavier have 4 apples* のような not が含まれる文、または含まれない文を追加し、GSM8K-pos/neg として同様の実験を行った。

文位置バイアス GSM8K の各問題の先頭・末尾に新たな文を追加したデータセットをそれぞれ GSM8K-head/tail として作成し、同様の実験を行った。

6.2 結果

結果は表 3 のようになった。いずれの表層の手がかりにおいても、手がかりに合致する文のほうが推論過程で言及される可能性が高く、多かれ少なか

表 3 解答を得るために必要のない文を 1 文追加した際に、LLM が CoT 中にその文を参照した確率. 質問文との重なりが大きい文、肯定的な文や文頭に存在する文の参照確率は高い

data type	model	GPT-3.5-turbo
GSM8K-overlap		59.5%
GSM8K-not-overlap		39.2%
GSM8K-pos		27.0%
GSM8K-neg		24.3%
GSM8K-head		39.2%
GSM8K-tail		27.0%

表 4 言語モデルが最短経路を出力しないようにバイアスを与えた形式言語データセット. 最短経路上の推論過程には質問文とは異なる主語を使用.

Context: **Peggy** has 5 apples. **Walter** has 2 more apples than Peggy. **Judy**'s mother has 3 less apples than Peggy. ...
Question: How many apples does **Judy** have?
Reasoning Processes: **Peggy** has 5 apples. **Walter** has 2 more apples than Peggy. ...

れ、探索時にこれらの手がかりを活用する傾向にあることが分かる。特に、類似文選択によるバイアスが強いことも分かる。

7 実験 2：探索戦略の動的変化

次に、これらの表層の手がかりが推論過程全体のうち、特にどのタイミング（1 ステップ目など）で特に用いられやすいかを調査する。この実験では、類似文選択を表層的な手がかりとする文を、最短経路上にない推論過程の出力を促すディストラクタとして含め、形式言語データを用意した。データはそれ以外のバイアスが生じないように統制されており、このデータ上で最短経路に合致する前提を選択できた場合は、最短経路選択戦略を実現できていると積極的に言える。表 4 にデータの例を示す。形式言語データセットは解の最短経路の長さが 3 になるように作成した。また、形式言語データを 100 例作成し、モデルは GPT-3.5-turbo と GPT-4 を使用した。

このようなデータ上で推論における各ステップにおいて、入力文中に表層的な特徴を有する文（ディストラクタ）を加え、次の推論ステップにおいて、モデルが最短経路上の前提を選択（出力）するか、ディストラクタを選択（出力）するかを測定した。なお推論ステップ n ($n > 1$) の設定の場合、 $n - 1$ ステップ目までの推論過程としては、最短経路における最初 n ステップをモデルに与えている。

図 3 に結果を示す。緑の線が、最短経路上の前提を出力した割合であり、オレンジの線がディストラ

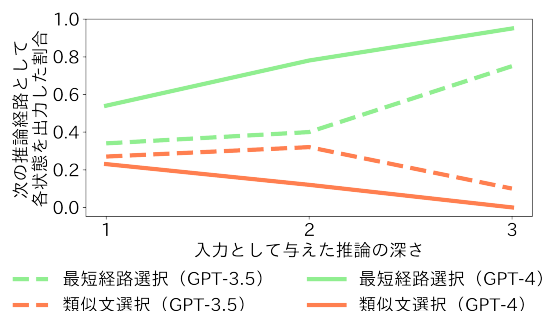


図 3 最短経路の長さが 3 の形式言語データセットを用意。緑：言語モデルに最短経路の 0 2 ステップまでを与えた時に推論過程 z_i として $[z_1^*, z_2^*, z_3^*]$ を出力した割合。オレンジ：言語モデルが論理的に妥当だが最短経路に含まれない推論過程 z_i を出力した割合

クタを選択した割合である。緑とオレンジの線に当てはまらない事例としては推論過程候補に存在しない、論理的に誤った出力が挙げられる。結果として、GPT-3.5, GPT-4 共に推論が進むにつれて表層的な手がかりに依拠した出力は起こりにくくなり、最短経路に合致する出力が増加することがわかった。したがって、推論の初期段階では広い探索空間から解答へと近づく最短経路を判断できず表層的な手がかりを用い、探索空間が狭まるにつれて最短経路を発見できることを示唆する。また、GPT-4 は GPT-3.5 と比較して、最短経路の選択が広い探索空間であっても高い確率で最短経路選択を行う能力を保持している事が分かった。これは、モデルサイズの増加に伴いより効率的な探索の実現可能性を示唆している。

8 おわりに

本研究では、大規模言語モデルの段階的推論において、探索に用いる手がかりが動的に変化することを明らかにした。具体的には、思考連鎖推論の初期段階では文の位置や質問文との類似度などの表層的な手がかりを用いた探索を行い、推論が進むにつれて最短経路を暗に計算する効率的な探索を行っている。本実験ではモデルとして GPT 系列のモデルのみを採用したが、今後はより多くのモデルを対象にしていきたい。また、本論文で示した結果はあくまでモデルの振る舞いから推察されるものであり、モデル内部においてどのような処理がされているかを直接確かめられたわけではない。今後、本論文で示したような、大規模言語モデルが行なっている探索戦略の変化を内部表現に着目して検証していくことも目指す。

謝辞

本研究は、JST CREST JPMJCR20D2 及び JSPS 科研費 JP21K21343 及び理化学研究所の基礎科学特別研究員制度の支援を受けたものである。

参考文献

- [1] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, **Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual**, 2020.
- [2] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. Llama: Open and efficient foundation language models. **CoRR**, Vol. abs/2302.13971, , 2023.
- [3] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In **NeurIPS**, 2022.
- [4] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. **CoRR**, Vol. abs/2305.10601, , 2023.
- [5] Bilgehan Sel, Ahmad Al-Tawaha, Vanshaj Khattar, Lu Wang, Ruoxi Jia, and Ming Jin. Algorithm of thoughts: Enhancing exploration of ideas in large language models. **CoRR**, Vol. abs/2308.10379, , 2023.
- [6] Herbert A. Simon. Search and reasoning in problem solving. **Artif. Intell.**, Vol. 21, No. 1-2, pp. 7–29, 1983.
- [7] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. In **NeurIPS**, 2022.
- [8] Andrew K. Lampinen, Ishita Dasgupta, Stephanie C. Y. Chan, Kory W. Mathewson, Mh Tessler, Antonia Creswell, James L. McClelland, Jane Wang, and Felix Hill. Can language models learn from explanations in context? In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022**, pp. 537–563. Association for Computational Linguistics, 2022.
- [9] Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah D. Goodman. Star: Bootstrapping reasoning with reasoning. In **NeurIPS**, 2022.
- [10] Mengnan Du, Fengxiang He, Na Zou, Dacheng Tao, and Xia Hu. Shortcut learning of large language models in natural language understanding: A survey. **CoRR**, Vol. abs/2208.11857, , 2022.
- [11] Yuxuan Lai, Chen Zhang, Yansong Feng, Quzhe Huang, and Dongyan Zhao. Why machine reading comprehension models learn shortcuts? In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, **Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021**, Vol. ACL/IJCNLP 2021 of **Findings of ACL**, pp. 989–1002. Association for Computational Linguistics, 2021.
- [12] Priyanka Sen and Amir Saffari. What do models learn from question answering datasets? In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020**, pp. 2429–2438. Association for Computational Linguistics, 2020.
- [13] Mengnan Du, Varun Manjunatha, Rajiv Jain, Ruchi Deshpande, Franck Dernoncourt, Jiuxiang Gu, Tong Sun, and Xia Hu. Towards interpreting and mitigating shortcut learning behavior of NLU models. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tür, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou, editors, **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021**, pp. 915–929. Association for Computational Linguistics, 2021.
- [14] Timothy Niven and Hung-Yu Kao. Probing neural network comprehension of natural language arguments. In Anna Korhonen, David R. Traum, and Lluís Màrquez, editors, **Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers**, pp. 4658–4664. Association for Computational Linguistics, 2019.
- [15] Miyoung Ko, Jinhyuk Lee, Hyunjae Kim, Gangwoo Kim, and Jaewoo Kang. Look at the first sentence: Position bias in question answering. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020**, pp. 1109–1121. Association for Computational Linguistics, 2020.

A GSM8K の拡張データセットの作成方法

GSM8K の各拡張データセットを作成するため、まず初めに以下のような前処理を行う。

- GSM8K の各問題文に対して、「人名または人称代名詞が含まれている」かつ「数値が含まれている」文を抽出し、各問題文に追加する文の候補として保存する
- 「問題文中に登場する人物が 1 人ではない」または「人名または人称代名詞が含まれていない」または「各問題文に追加する文候補がない」問題を削除する。
- すべての人称代名詞をその人称代名詞が指している人物名に変更
- 問題文に追加する文候補の数値を変更

続いて、各拡張データセット毎に、保存済みの文候補から各問題に文を 1 つ追加する処理を行う。

A.1 GSM8K-over/not-over

GSM8K-over, GSM8K-not-over では GSM8K の各問題に追加する文の主語を以下のように変化させた。

- over: 質問文中の人物名 + 「s mother"|"'"s father"|"'"s son"|"'"s neighborhood」
- not-over: 質問文中に登場しない人物名

この処理によって、新たに追加した文と質問文の重なりが相対的に大きい/小さいデータセットとして、GSM8K-over/GSM8K-not-over が作成できる。

B プロンプト

実験 2 : 探索戦略の動的変化では表 5 のプロンプトを使用した。

表 5 形式言語データセットを解く際の入力プロンプト

Answer the context question according to the following example.

Context: Walter has 1 apples. Ursula has 3 more apples than Walter. Victor has 3 more apples than Ursula. Quentin has 2 more apples than Ursula. Nancy has 3 more apples than Walter. Zoe has 3 more apples than Nancy. Heidi has 3 more apples than Nancy. Carol has 4 apples. Xavier has 3 more apples than Carol. Peggy has 4 more apples than Xavier. Dave has 3 more apples than Xavier. Bob has 1 more apples than Carol. Alice has 3 more apples than Bob. Sybil has 2 more apples than Bob.

Question: How many apples does Dave have?

Answer:

Carol has 4 apples, and Xavier has 3 more apples than Carol. So, Xavier has 7 apples.

Xavier has 7 apples, and Dave has 3 more apples than Xavier. So, Dave has 10 apples.

The final answer is 10.

Context: Context: Zoe has 6 more apples than Larry. Ivan has 4 more apples than Nancy. Olivia has 3 apples. Zoe has 3 less apples than Sybil. Sybil has -1 apples. Nancy has 0 apples. Carol has 5 less apples than Olivia.

Question: How many apples does Zoe have?

Answer:

Sybil has -1 apples, and Zoe has 3 less apples than Sybil. So, Zoe has -4 apples.

The final answer is -4.

Context: Context: Zoe has 10 more apples than Dave. Eve has 2 apples. Dave has 3 more apples than Eve. Yvonne has 3 more apples than Quentin. Yvonne has 3 less apples than Zoe. Zoe has 3 more apples than Grace. Trent has 3 more apples than Zoe. Ivan has 3 apples. Ursula has 3 more apples than Zoe. Grace has 3 apples. Zoe has 3 more apples than Ivan.

Question: How many apples does Yvonne have?

Answer: Eve has 2 apples, and Dave has 3 more apples than Eve. So, Dave has 5 apples.

Dave has 5 apples, and Zoe has 10 more apples than Dave. So, Zoe has 15 apples.

Zoe has 15 apples, and Yvonne has 3 less apples than Zoe. So, Yvonne has 12 apples.

The final answer is 12. **Context:** Victor has 1 more apples than Mallory. Judy has 3 less apples than Peggy. Rob has -2 more apples than Eve. Peggy has -5 apples. Walter has 3 more apples than Peggy. Eve has 3 apples. Peggy has 3 more apples than Walter. Mallory has 3 more apples than Eve.

Question: How many apples does Victor have?

Answer: Eve has 3 apples, and Mallory has 3 more apples than Eve. So, Mallory has 6 apples.

Mallory has 6 apples, and Victor has 1 more apples than Mallory. So, Victor has 7 apples.

The final answer is 7.