TOWARD A REASONING CURRICULUM FOR BRAIN-TRAINED FOUNDATION MODELS

Maël Donoso
Ouroboros Neurotechnologies
Lausanne, Switzerland
mael.donoso@ouroboros-neurotechnologies.com

ABSTRACT

The processes of human reasoning can be predicted by computational models of human cognition, and have been associated with the neural signals recorded from several brain regions, particularly in the prefrontal cortex. Theoretically, these neural signals could be directly leveraged for training the reasoning capabilities of foundation models, in line with the emerging approach of using neuroimaging data to improve the performance of artificial intelligence systems. Practically, following this new strategy would require the development of a reasoning curriculum, a progression from simple and narrow experimental settings to more complex and open-ended reasoning experiments. This reasoning curriculum could provide valuable insights for brain-trained foundation models, and could eventually be integrated into human-written supervised fine-tuning demonstrations, which are already part of the post-training strategies of foundation models.

1 HUMAN REASONING

Human reasoning engages a complex network of brain regions, with the prefrontal cortex playing a prominent role. In particular, the ventromedial prefrontal cortex (vmPFC) encodes the value of the elements present in the environment (Lebreton et al., 2009), as well as more abstract estimates such as the value of the chosen action (Wunderlich et al., 2009), the value of the current strategy (Kolling et al., 2012), and the degree of reliability associated with this strategy (De Martino et al., 2013). Meanwhile, the dorsolateral prefrontal cortex (DLPFC) implements a hierarchy of increasingly abstract executive processes, controlling behavior based on stimuli, context, and instructions (Koechlin et al., 2003), and allowing for the acquisition of abstract rules (Badre et al., 2010). Human reasoning also involves the anterior cingulate cortex (ACC), which encodes the volatility of the environment (Behrens et al., 2007) and the value of exploratory behavior (Kolling et al., 2012), as well as the frontopolar cortex (FPC), which encodes the reliabilities of alternative options (Boorman et al., 2009). Finally, at the subcortical level, the ventral striatum plays a fundamental role in reward prediction (O'Doherty et al., 2004) and motivation (Pessiglione et al., 2007). Together with additional brain regions such as the posterior parietal cortex, this network processes rewards, estimates the value of elements, actions, and strategies, monitors the reliabilities of current and alternative options, and implements decision-making based on a hierarchy of rules.

To provide a realistic view of human reasoning, a computational model must explain our ability to select behavioral strategies in uncertain, changing, and open-ended environments, while relying on biologically plausible mechanisms. One such model (Collins & Koechlin, 2012) proposes that the current strategy, called the actor, is continuously adapted using reinforcement learning processes, while the absolute reliability of several monitored strategies is simultaneously estimated using Bayesian inference. In this model, exploitation periods correspond to the time steps where the actor is considered reliable enough. When the actor becomes unreliable, perhaps because a change in the environment has rendered the current policy obsolete, the model switches to an exploration period, and a new acting strategy, called the probe, is implemented as a weighted mixture of all previously used strategies. The return to exploitation is controlled by a hypothesis testing mechanism, with two possible outcomes: either the probe is successfully adapted and becomes reliable, in which case it is confirmed; or one of the few alternative strategies monitored in working memory reaches the reliability threshold, in which case the probe is rejected and this alternative strategy becomes

the new actor. Here, the details of the model, which has been tested and validated by behavioral experiments, are less important than the general prediction that human reasoning requires an estimation of the values and reliabilities of behavioral strategies, as well as the implementation of three algorithmic transitions: switch, confirmation, and rejection.

Importantly, it has been demonstrated that a single reasoning experiment, conducted using functional magnetic resonance imaging (fMRI), can simultaneously reveal the foundations of all these computational processes in the human brain (Donoso et al., 2014), uncovering a neural system consistent with previous neuroscience literature. In particular, the vmPFC, which is known to encode the value of the current strategy, correlates with the reliability of the actor; the DLPFC, which controls behavior based on abstract rules, correlates with the rejection transition, where an alternative strategy is recovered; the ACC, which encodes the value of exploratory behavior, correlates with the switch transition, where an exploration period is initiated; the FPC, which encodes the reliabilities of alternative options, correlates indeed with the monitoring of alternative strategies; and the ventral striatum, which is engaged in reward prediction and motivation, correlates with the confirmation transition, where a probe is validated as a new successful strategy. The fact that all these neural signals can be simultaneously uncovered within a single reasoning experiment is nontrivial, since such signals could provide direct insights for guiding the reasoning capabilities of artificial intelligence (AI) models, as we will now discuss.

2 Artificial Intelligence Reasoning

Every example of human reasoning included in the training corpus of foundation models, such as the written reasoning examples included in the training corpus of large language models (LLMs), is ultimately the product of human brains. Although inference techniques such as chain-of-thought (CoT) prompting (Wei et al., 2023) can improve the ability of foundation models to solve complex tasks in a manner that approximates human cognition, the performance of these models remains limited by the absence of a cognitive control architecture comparable to the human prefrontal cortex (Russin et al., 2020). As a result, foundation models often perform poorly on reasoning tasks outside their training data distribution (Stechly et al., 2025). Furthermore, although several approaches have been proposed to overcome this challenge, including techniques such as tree of thoughts (Yao et al., 2023) and graph of thoughts (Besta et al., 2024), these approaches typically result in greater computational costs, and may still underperform direct answering across various benchmarks (Zheng et al., 2025). It is natural to think that neuroscience-inspired architectures, more grounded in our knowledge of human cognition, could open new horizons for improving AI reasoning—but there may also be another, more direct way. A new strategy is currently emerging at the intersection of neuroscience and AI: training foundation models directly on human brain data, using any learning, regularization, fine-tuning, or alignment method that leverages neuroimaging datasets to improve the performance of a model. This new strategy could be particularly relevant for bridging the gap between human cognition and AI reasoning, by directly accessing the neural signals recorded during human reasoning experiments.

Existing research has already demonstrated that neuroimaging data, acquired using techniques such as fMRI, electroencephalography (EEG), or magnetoencephalography (MEG), can improve the performance of classical machine learning, deep learning, or reinforcement learning models for a variety of tasks. In the visual domain, studies have demonstrated that fMRI data can be leveraged to improve classical machine learning models (Fong et al., 2018), and EEG data to improve deep learning models (Spampinato et al., 2017) for object recognition. fMRI data can also be used to guide deep learning models in the auditory domain (Freteault et al., 2025), and EEG data to guide reinforcement learning models for robotic movements (Iturrate et al., 2010). It has also been demonstrated that fMRI and MEG data can serve to align language models with human semantic representations (Schwartz et al., 2019), and that fMRI data can be used to improve the performance of speech language models on downstream semantic tasks (Moussa et al., 2025). Together, these results suggest that neuroimaging data can provide insights into human cognition that would not be accessible through observable actions alone, such as text or speech. As a result, neuroimaging datasets could emerge as a valuable resource for complementing the classical training corpus of foundation models, at a time when data scaling and computing power scaling may start to yield diminishing returns in terms of model performance.

3 BRAIN-TRAINED FOUNDATION MODELS

Two high-level methods have been proposed to prioritize the use of neuroimaging data, which is relatively scarce and expensive compared with text or speech data, for strategically chosen, high-value steps in foundation model training: reinforcement learning from human brain (RLHB) and chain of thought from human brain (CoTHB) (Donoso, 2026). Such methods could provide a practical path toward *brain-trained foundation models*, combining classical training data with the neural signals extracted from a series of brain regions of interest. Brain-trained foundation models should not be confused with the large models developed for neuroscience and neurotechnology, sometimes known as *brain foundation models*. Whereas brain foundation models are used to predict neural patterns within specific neuroimaging modalities, brain-trained foundation models would leverage neuroimaging data to improve the performance of general-purpose foundation models, such as LLMs, on a wide range of tasks. Of the two methods we just mentioned, RLHB could be the most straightforward to implement, as it is natural to envision an extension of reinforcement learning from human feedback (RLHF) (Christiano et al., 2017) where the neural signals associated with reward, value, confidence, surprise, effort, or conflict would be leveraged for a deeper representation of human preferences. CoTHB could be more difficult to realize, but it has the potential to significantly improve the reasoning capabilities of foundation models, by reaching beyond the surface-level patterns that can be extracted from written or spoken reasoning examples.

Theoretically, many different CoTHB algorithms could be envisioned, directly based on the computational processes of the human brain, and utilizing the neural signals extracted from the regions of interest to guide the chain of thought, or any form of thinking path, of foundation models. For example, as long as the reliability of the current strategy (inferred from vmPFC activity) stays above a certain threshold, the model could continue its current thinking path; otherwise, it could engage in a new, exploratory path. The reliabilities of the alternative options (FPC) could serve as an indicator for monitoring a specific number of alternative thinking branches, parallel to the current thinking path. The three algorithmic transitions we mentioned previously, namely the switch (ACC), confirmation (ventral striatum), and rejection (DLPFC) transitions, could serve as markers for abandoning the current thinking path, validating an exploratory path, or recovering a previous thinking path, respectively. Regardless of the specific algorithms, leveraging the neural signals associated with human reasoning could represent a promising solution for guiding the thinking strategies of foundation models, in a manner more consistent with our knowledge of human cognition.

Practically, however, methods such as CoTHB would depend on a critical factor: the successful generalization from simple and narrow reasoning experiments to the complex and open-ended reasoning problems that we could expect foundation models to solve. Existing research on training AI models using human brain data has often focused on perceptual tasks, such as object recognition or audio classification, or on simple executive tasks, such as the generation of reward and error signals for robotics. In both situations, there is a relatively straightforward correspondence between the task performed by the human brain and the task for which the AI model is trained. At a different level, methods such as RLHB seem realistic because RLHF sessions are already routinely organized for foundation model post-training, suggesting that the addition of at least some neuroimaging modalities, such as EEG, could be a feasible extension of current post-training pipelines. By contrast, methods such as CoTHB could be more challenging, regardless of the neuroimaging technique, because current human reasoning experiments are often conducted using highly constrained, non-naturalistic experimental designs. For example, the reasoning experiment that simultaneously revealed the foundations of multiple reasoning processes in the human brain (Donoso et al., 2014) relied on a Mastermind-like game optimized for fMRI acquisition, where the primary stimuli were limited to three visually presented digits, and the actions that the subject could choose were limited to four possible buttons. Therefore, while the prefrontal cortex and related brain regions may hold precisely the type of neural signals that would be needed to improve the reasoning capabilities of foundation models, leveraging the insights from reasoning experiments may require the development of new experimental designs, allowing for the extraction of the neural signals of interest in more open-ended and naturalistic settings.

A related challenge is the fact that neuroimaging data can only be acquired over short periods, whereas foundation models must often reason over problems that would take much longer for humans to solve. Still, since long reasoning problems can often be decomposed into shorter steps, it seems realistic to think that brain-trained foundation models could be able to extrapolate at least

some patterns from relatively short reasoning experiments, assuming that these experiments are designed to allow for meaningful generalization. Regarding the costs, while methods such as CoTHB would probably require the acquisition of new neuroimaging datasets, these investments may be justified by the strategic importance of this research direction, and the expected benefits of brain-trained foundation models. Furthermore, existing research on training AI models using human brain data has reported performance improvements even with a limited number of subjects, despite the noise, individual variation, and other challenging characteristics of neuroimaging data, suggesting that the extraction of the neural signals of interest could be done relatively efficiently. Nevertheless, to ensure that these new neuroimaging datasets can be utilized as a valuable resource for guiding the reasoning capabilities of foundation models, it seems necessary to devise at least a high-level plan for progressing toward more open-ended and naturalistic reasoning tasks.

4 Toward a Reasoning Curriculum

This naturally leads to the idea of a *reasoning curriculum*: a progression of experimental paradigms that would keep the computational structure of human reasoning intact, while gradually increasing the input space and output space of the reasoning problem. In other words, in parallel with the theoretical exploration of possible CoTHB algorithms, we should envision a succession of experimental designs where we assume that reliability monitoring, hypothesis testing, and other computational processes of human reasoning would remain tractable and measurable by neuroimaging techniques, but where both the stimuli presented to the subjects and the set of available actions would become gradually more diverse.

A high-level plan for this reasoning curriculum could involve a three-step progression, as illustrated in Figure 1. In the first step, corresponding to the current situation, reasoning experiments are mostly based on non-verbal, non-naturalistic experimental designs. In the second step, we could try to replicate previous results using constrained verbal reasoning tasks, for example by devising new Mastermind-like games governed by simple semantic rules or contexts; in such settings, subjects could be asked to complete fragments of text or answer simple questions, either by selecting, saying, or writing simple verbal outputs. In the third step, we could try to further expand verbal reasoning into more open-ended and naturalistic tasks, where the semantic rules or contexts could become more complex; in such settings, subjects could be asked to solve realistic problems, either orally or in writing, therefore providing more elaborate verbal outputs. Naturally, this high-level, three-step plan provides only an intuitive roadmap. There may be multiple intermediate steps throughout the process, but the testable hypothesis remains that the degrees of freedom of the reasoning problem could be gradually extended, while preserving our capacity to monitor the neural signals of interest associated with human reasoning.

Fundamentally, starting with the second step (constrained verbal reasoning), the use of verbal inputs and outputs could render the neural signals actionable in the context of foundation model training, for example LLM training. As a result, the neuroimaging data acquired in the second and third steps could naturally form the basis of an actual curriculum learning strategy (Bengio et al., 2009), where neural signals and their corresponding semantic data could be meaningfully ordered from simple to complex tasks and environments, potentially improving training and generalization. In this sense, the acquisition of new, carefully designed neuroimaging datasets could directly translate into improvements in the reasoning capabilities of foundation models. Remarkably, if the computational processes of human reasoning can indeed be monitored in verbal settings of increasing complexity, the reasoning experiments of the third step (open-ended verbal reasoning) could eventually converge toward a form very similar to human-written supervised fine-tuning (SFT) demonstrations (Ouyang et al., 2022), where human experts naturally engage in complex reasoning by providing high-quality answers to various questions. Therefore, whereas the acquisition of dedicated neuroimaging datasets could be necessary for the second step, it is conceivable that the third step could eventually be more seamlessly integrated into the SFT sessions that are already organized for foundation model post-training. In such a scenario, while RLHB may be defined as the addition of neuroimaging to RLHF, this particular form of CoTHB might be symmetrically defined as the addition of neuroimaging to SFT demonstrations.

Looking forward, if neural interfaces become more widely adopted in the future, the computational processes of human reasoning might even be monitored at scale in a diversity of real-life situations.

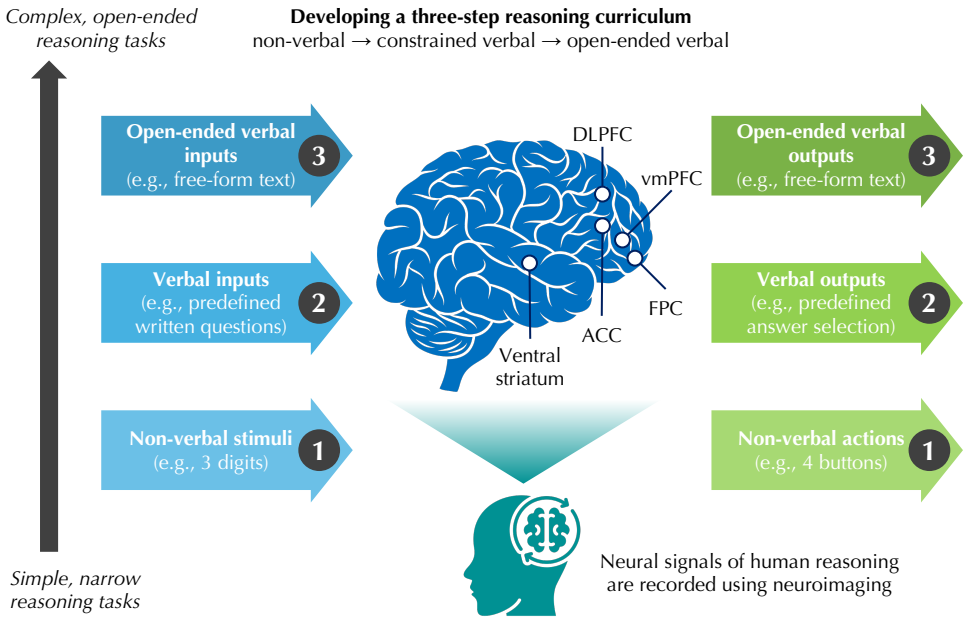

Figure 1: **Developing a three-step reasoning curriculum**. The neural signals associated with human reasoning (e.g., vmPFC, DLPFC, ACC, FPC, and ventral striatum activity) are recorded using neuroimaging, and are hypothesized to remain invariant across task and environment complexities. The three steps represent a progression from simple, narrow reasoning tasks to complex, open-ended reasoning tasks. *Step 1*: current experimental paradigms often focus on non-verbal reasoning, with a limited set of primary stimuli (e.g., 3 digits) and possible actions (e.g., 4 buttons). *Step 2*: constrained verbal reasoning paradigms would keep the computational structure of human reasoning intact, while increasing the input space (e.g., predefined written questions) and output space (e.g., predefined answer selection), and rendering the neural signals actionable in the context of foundation model training. *Step 3*: open-ended verbal reasoning paradigms would further increase the degrees of freedom for both inputs and outputs (e.g., free-form text), and potentially converge toward a form very similar to human-written SFT demonstrations.

This would require that the neural signals identified using sophisticated and expensive neuroimaging techniques, such as fMRI or MEG, can be captured using neuroimaging modalities more compatible with wearable devices, such as EEG. While this development remains hypothetical, it could be considered a potential fourth step in the reasoning curriculum.

Large Language Model Usage LLMs were used at several stages of this research to: (1) brainstorm scientific ideas to a limited extent, and (2) improve the readability of the text through English correction. All final decisions regarding the manuscript content were made by the author.

References

David Badre, Andrew S. Kayser, and Mark D'Esposito. Frontal cortex and the discovery of abstract action rules. *Neuron*, 66(2):315–326, 2010. doi: 10.1016/j.neuron.2010.03.025.

Timothy E. J. Behrens, Mark W. Woolrich, Mark E. Walton, and Matthew F. S. Rushworth. Learning the value of information in an uncertain world. *Nature Neuroscience*, 10(9):1214–1221, 2007. doi: 10.1038/nn1954.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 41–48, 2009. doi: 10.1145/1553374.1553380.

Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michal Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, and Torsten Hoefler. Graph of Thoughts: Solving Elaborate Problems with Large Language Models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(16):17682–17690, 2024. doi: 10.1609/aaai.v38i16.29720.

Erie D. Boorman, Timothy E. J. Behrens, Mark W. Woolrich, and Matthew F. S. Rushworth. How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. *Neuron*, 62(5):733–743, 2009. doi: 10.1016/j.neuron.2009.05.014.

Paul F. Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep Reinforcement Learning from Human Preferences. *Advances in Neural Information Processing Systems*, 30, 2017.

Anne Collins and Etienne Koechlin. Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biology*, 10(3):e1001293, 2012. doi: 10.1371/journal.pbio.1001293.

Benedetto De Martino, Stephen M. Fleming, Neil Garrett, and Raymond J. Dolan. Confidence in value-based choice. *Nature Neuroscience*, 16(1):105–110, 2013. doi: 10.1038/nn.3279.

Maël Donoso. A New Strategy for Artificial Intelligence: Training Foundation Models Directly on Human Brain Data. *arXiv*, 2026. doi: 10.48550/arXiv.2601.12053.

Maël Donoso, Anne G. E. Collins, and Etienne Koechlin. Foundations of human reasoning in the prefrontal cortex. *Science*, 344(6191):1481–1486, 2014. doi: 10.1126/science.1252254.

Ruth C. Fong, Walter J. Scheirer, and David D. Cox. Using human brain activity to guide machine learning. *Scientific Reports*, 8(1):5397, 2018. doi: 10.1038/s41598-018-23618-6.

Maëlle Freteault, Maximilien Le Clei, Loic Tetrel, Lune Bellec, and Nicolas Farrugia. Alignment of auditory artificial networks with massive individual fMRI brain data leads to generalisable improvements in brain encoding and downstream tasks. *Imaging Neuroscience*, 3:imag_a_00525, 2025. doi: 10.1162/imag_a_00525.

Iñaki Iturrate, Luis Montesano, and Javier Minguez. Robot reinforcement learning using EEG-based reward signals. *2010 IEEE International Conference on Robotics and Automation*, pp. 4822–4829, 2010. doi: 10.1109/ROBOT.2010.5509734.

Etienne Koechlin, Chrystèle Ody, and Frédérique Kouneiher. The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*, 302(5648):1181–1185, 2003. doi: 10.1126/science.1088545.

Nils Kolling, Timothy E. J. Behrens, Rogier B. Mars, and Matthew F. S. Rushworth. Neural Mechanisms of Foraging. *Science*, 336(6077):95–98, 2012. doi: 10.1126/science.1216930.

Maël Lebreton, Soledad Jorge, Vincent Michel, Bertrand Thirion, and Mathias Pessiglione. An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. *Neuron*, 64(3):431–439, 2009. doi: 10.1016/j.neuron.2009.09.040.

Omer Moussa, Dietrich Klakow, and Mariya Toneva. Improving Semantic Understanding in Speech Language Models via Brain-tuning. *arXiv*, 2025. doi: 10.48550/arXiv.2410.09230.

John O'Doherty, Peter Dayan, Johannes Schultz, Ralf Deichmann, Karl Friston, and Raymond J. Dolan. Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. *Science*, 304(5669):452–454, 2004. doi: 10.1126/science.1094285.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F. Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.

Mathias Pessiglione, Liane Schmidt, Bogdan Draganski, Raffael Kalisch, Hakwan Lau, Ray J. Dolan, and Chris D. Frith. How the Brain Translates Money into Force: A Neuroimaging Study of Subliminal Motivation. *Science*, 316(5826):904–906, 2007. doi: 10.1126/science.1140459.

Jacob Russin, Randall C. O'Reilly, and Yoshua Bengio. Deep Learning Needs a Prefrontal Cortex. *Work Bridging AI Cogn Sci*, 2020.

Dan Schwartz, Mariya Toneva, and Leila Wehbe. Inducing brain-relevant bias in natural language processing models. *Advances in Neural Information Processing Systems*, 32, 2019.

Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Nasim Souly, and Mubarak Shah. Deep Learning Human Mind for Automated Visual Classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6809–6817, 2017. doi: 10.1109/CVPR.2017.479.

Kaya Stechly, Karthik Valmeekam, and Subbarao Kambhampati. Chain of Thoughtlessness? An Analysis of CoT in Planning. *arXiv*, 2025. doi: 10.48550/arXiv.2405.04776.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *arXiv*, 2023. doi: 10.48550/arXiv.2201.11903.

Klaus Wunderlich, Antonio Rangel, and John P. O'Doherty. Neural computations underlying action-based decision making in the human brain. *Proceedings of the National Academy of Sciences*, 106 (40):17199–17204, 2009. doi: 10.1073/pnas.0901077106.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of Thoughts: Deliberate Problem Solving with Large Language Models. *Advances in Neural Information Processing Systems*, 36:11809–11822, 2023.

Tianshi Zheng, Yixiang Chen, Chengxi Li, Chunyang Li, Qing Zong, Haochen Shi, Baixuan Xu, Yangqiu Song, Ginny Y. Wong, and Simon See. The Curse of CoT: On the Limitations of Chain-of-Thought in In-Context Learning. *arXiv*, 2025. doi: 10.48550/arXiv.2504.05081.