NATHANIEL GAN[1]

# IMAGINATION-BASED ARTIFICIAL INTELLIGENCE

**Abstract**
This paper highlights imagination-based artificial intelligence as a new and promising approach to artificial intelligence (AI). The characteristic features of imagination-based AI are a framing of target tasks in terms of affordances and an identification of those affordances using internal simulations. Systems incorporating the imagination-based approach have attained comparable performance levels to deep-learning-based models while having lower data requirements and greater explainability. Work remains to be done to develop the breadth of applicability of the imagination-based approach, and this class of AI systems raises philosophically interesting issues regarding the possibility of imagination in artificial systems.
**Key words**: artificial intelligence; imagination; simulations; affordances; deep learning

**Inteligência artificial baseada na imaginação**

**Resumo**
Este artigo destaca a inteligência artificial baseada na imaginação como uma nova e promissora aproximação à inteligência artificial (AI). As características da IA baseada na imaginação são a definição das tarefas-alvo em termos de *affordances* e a identificação dessas *affordances* pelo uso de simulações internas. Os sistemas que incorporam a aproximação baseada na imaginação atingiram níveis de desempenho comparáveis aos

[1] University of Singapore


DOI: https://doi.org/10.21747/21836892/fil42a13

modelos baseados em *deep-learning* apesar de terem requisitos de dados mais baixos e uma maior compreensibilidatde. Ainda há trabalho a fazer para desenvolver a amplitude de aplicabilidade das propostas baseadas na imaginação, e esta classe de sistemas de IA levanta questões filosóficas interessantes acerca da possibilidade da imaginação em sistemas artificiais.

**Palavras-chave**: inteligência artificial; imaginação; simulações; affordances; deep learning

Deep learning currently represents one of the most popular approaches to artificial intelligence (AI), having found a wide range of applications to autonomous vehicles, chatbots, and game-playing engines, among other things. This interest in deep learning is primarily motivated by its advantages over traditional machine learning models, compared to which deep-learning-based models are able to process larger volumes of data more efficiently to perform at higher levels. But deep learning also has its drawbacks. Neural networks are known to require large amounts of training data and be computationally expensive. It is also often challenging to interpret trained models or explain their decisions. So despite the current popularity and success of deep learning, it might be worth exploring other approaches to AI.

This paper aims to highlight a recent and lesser-known approach to AI that presents a viable alternative to deep learning: *imagination-based* AI[2]. It will be seen that imagination-based AI shares some of the advantages that deep learning has over traditional machine learning. At the same time, imagination-based AI mitigates the weaknesses of deep learning pertaining to data and computational requirements, and offers greater interpretability. As an area for research, furthermore, imagination-based AI provides potentially fruitful avenues for technical and philosophical exploration.

In what is to come, §1 introduces the imagination-based approach to AI. §2 makes a case for the viability of this approach by arguing that it shares some strengths of deep learning while mitigating some of its weaknesses. §§3–4 highlight potential areas for technical and philosophical research.

## 1. Imagination-based AI

The imagination-based approach to AI has two distinctive characteristics. The first concerns the conceptualisation of the target task. According to some popular views in ecological psychology, animals learn to function in their environment by perceiving possible ways in which they can interact with objects around them—James Gibson[3] called these possible interactions *affordances* [4]. Imagination-based

[2] The philosophical appropriateness of this term will be considered in §4.

[3] J. J. Gibson. *The Perception of the Visual World*. Houghton Mifflin, Boston 1950 and *The Ecological Approach to Visual Perception*. Routledge, New York 1979.

[4] Views differ in the literature over how precisely to understand the nature of affordances and their role in natural selection—see I. Chong, I. & R. W. Proctor, «On the evolution of a radical concept: Affordances according to Gibson and their subsequent use and development» in *Perspectives on Psychological Science, 15*(1) (2019)117–132.

AI systems adapt this idea by framing their tasks in terms of affordances. When attempting to identify a chair in its environment, for instance, an imagination-based system might seek to detect objects that afford sitting, and take this affordance as a key indicator for whether a given object is a chair. This is in contrast to deep learning models, which typically frame object-detection tasks in terms of pattern-matching for visual features.

With a task framed in terms of affordances, a second distinctive feature of imagination-based AI is that it attempts to detect these affordances using internal simulations. When attempting to determine if an object affords sitting, an imagination-based system might run simulations in which a human attempts to sit on the object, and deduce the presence of the sitting-affordance based on the results of the simulation. To be sure, the use of representations of world states to inform model decisions is already a known technique in AI systems[5]. The imagination-based approach develops this technique further by integrating the use of world models with affordance-detection methods.

An example of imagination-based AI is the chair-detection system introduced in Wu et al.[6] and implemented in Wu et al.[7] and Meng et al.[8] . This system uses an affordance-based definition of 'chair' as follows:

> An object which can be stably placed on a flat horizontal surface in such a way that a typical human is able to sit (adopt or rest in a posture in which the body is supported on the buttocks and thighs and the torso is more or less upright) stably above the ground[9].

Given an unidentified object, the system takes as input images of the target object from different perspectives (when implemented in an embodied system, the images are obtained using a camera attached to a robot). An internal three-dimensional simulation of the target object in its present orientation is then constructed.

Then, the system performs two series of simulations to determine, respectively, the orientations in which the object can be stably placed on the ground and whether a human can sit stably on the object in its stable orientations. In the first series of simulations, the system simulates the motion of the object under the usual physical laws when dropped from a height after some rotation in 3D space. An orientation is considered stable if, when dropped from that orientation, the rotational and

[5] D. Ha, & J. Schmidhuber, «World models» in *arXiv preprint* (2018) 1–21; Schmidhuber, J. «On learning to think: Algorithmic information theory for novel combinations of reinforcement learning controllers and recurrent neural world models». *arXiv preprint* (2015)1–36. Thanks to a reviewer for highlighting this point.

[6] H. Wu, D. Misra & G. S. Chirikjian, «Is that a chair? Imagining affordances using simulations of an articulated human body». *IEEE International Conference on Robotics and Automation*, Paris, 2020 :10.1109/ICRA40945.2020.9197384

[7] H. Wu, D. Misra & G. S. Chirikjian *Put the bear on the chair! Intelligent robot interaction with previously unseen chairs via robot imagination*. arXiv preprint (2022a):arXiv:2108.05539

[8] Meng, X., Wu, H., Ruan, S., & Chirikjian, G. S.. «Prepare the chair for the bear! Robot imagination of sitting affordance to reorient previously unseen chairs». *arXiv preprint* (2023): doi:arXiv:2306.11448v1

[9] H. Wu & G.S. Chirikjian, Can I pour into it? Robot imagining open containability affordance of previously unseen objects via physical simulations. *IEEE Robotics and Automation Letters, 6*(1) (2020) 271–278.

translational motion of the object is below a threshold (determined during training[10]). In the second series of simulations, the simulated object is oriented to a stable position, and a simulated humanoid is dropped onto it. The humanoid's final resting position is considered a sitting position if its height, joint angles, and contact with the object fall within predetermined thresholds. The target object is classified as a chair if for at least one orientation, the simulated humanoid finds a sitting position at least once.

Besides this example, the imagination-based approach has also been applied to object rearrangement tasks[11], and the identification of other objects like containers[12], frying pans[13], tools[14], and others[15].

## 2. Strengths of imagination-based AI

The imagination-based approach to AI is relatively unknown—whereas deep learning has found a wide range of applications, the use of imagination-based AI has hitherto not extended far beyond the examples cited above. Nevertheless, this section will argue, the imagination-based approach has significant strengths.

To begin with, the performance levels of imagination-based systems are comparable to those attained by deep learning models. One consideration typically raised in favour of deep learning is the improved performance of deep learning algorithms relative to traditional machine learning algorithms. In a 2017 study comparing various models aimed at predicting energy consumption, for instance, it was found that deep learning models surpass benchmarks set by models based on traditional machine learning algorithms such as linear regression, support vector machines, decision trees, and Gaussian processes[16]. Deep learning models have also been shown to perform at higher levels than traditional machine learning models at

[10] The system uses a training set of 30 objects from the chair class of the Princeton ModelNet40 dataset (Wu et al., 2015). The thresholds are chosen in such a way as to maximise sitting quality and accuracy across the training examples.

[11] Wu, H., Ye, J., Meng, X., Paxton, C., & Chirikjian, G. S. (2022b). *Transporters with visual foresight for solving unseen rearrangement tasks*. arXiv preprints.

[12] Wu, H., & Chirikjian, G. S. (2020). Can I pour into it? Robot imagining open containability affordance of previously unseen objects via physical simulations. *IEEE Robotics and Automation Letters, 6*(1), 271–278.

[13] Ho, S. B. (2023). Why is that a good or not a good frying pan? Knowledge representation for functions of objects and tools for design understanding, improvement, and generation. *arXiv*. doi:10.48550/arXiv.2303.06152

[14] Zhu, Y., Zhao, Y., & Zhu, S. (2015). Understanding Tools: Task-Oriented Object Modeling, Learning and Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 2855–2864. doi:10.1109/CVPR.2015.7298903

[15] Bar-Aviv, E., & Rivlin, E. (2016). Functional 3D Object Classification Using Simulation of Embodied Agent. *Proceedings of the British Machine Vision Conference*, 307–316.

[16] N. G. Paterakis, E. Mocanu, M. Gibescu, B. Stappers & W. Alst,. «Deep Learning Versus Traditional Machine Learning Methods for Aggregated Energy Demand Prediction in *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe* (2017) 1–6.

tasks like image recognition[17], natural language processing[18], fraud detection[19], and others.

Imagination-based AI shares this advantage of deep learning, having attained performance levels comparable to those attained by state-of-the-art deep learning models where the two approaches are applicable. The chair-detection system introduced in §1 attained 97% accuracy on a test set comprising 300 labelled images of randomly oriented objects, and an accuracy of 94.4% in physical experiments with previously unseen objects. In comparison, appearance-based models that use convolutional neural networks attained below 90% accuracy on the same test set[20]. Some other examples are given in the table below.

| **Task** | **Imagination-based systems** | **Deep learning models** |
|---|---|---|
| Chair detection | (Wu et al., 2022a; Wu et al., 2020): Accuracy = 97%<br><br>(Hinkle & Olson, 2013): Accuracy = 90% | (Kanezaki et al., 2018): Accuracy = 83.0%<br><br>(Su et al., 2018): Accuracy = 86.3% |
| Object rearrangement | (Wu et al., 2022b): Accuracy = 74.35% | (Seita et al., 2021): Accuracy = 48.47% |
| Container detection | (Wu & Chirikjian, 2020): Accuracy = 100%; Success rate = 98.18%<br><br>(Hinkle & Olson, 2013): Accuracy = 93% | (Do et al., 2018): Accuracy = 100%; Success rate = 91.30% |

The imagination-based approach also has an advantage over deep learning when it comes to data requirements. Deep learning models often require numerous training examples to reach their performance levels. The deep-learning-based chair-detection system in Kanezaki, Matsushita & Nishida [21] attained the performance level observed above after training on close to 10,000 examples, while the deep-learning-based container-detection system in Do et al**.** [22] was trained on 21,000 examples. The high data requirements of deep learning models make them computationally costly,

[17] Y. Lai, «A Comparison of Traditional Machine Learning and Deep Learning in Image Recognition», in *Journal of Physics: Conference Series, 1314*, (2019) 1–8.

[18] Li, H. «Deep learning for natural language processing: advantages and challenges», *National Science Review, 5*(1), (2018) 24–26.

[19] Gupta, S. *Deep learning vs. traditional machine learning algorithms used in credit card fraud detection.* (Masters thesis). National College of Ireland (2016).

[20] A. Kanezaki, Y. Matsushita, & Y. Nishida, «Rotationnet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints». *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018) 5010–5019; J. C. Su, M.Gadelha, R. Wang & S. Maji «A deeper look at 3d shape classifiers». *Second Workshop on 3D Reconstruction Meets Semantics* (2018).

[21] A. Kanezaki, Y. Matsushita & Y. Nishida «Rotationnet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints», in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5010–5019.

[22] T. Do, A. Nguyen, & , I. Reid «Affordancenet: An end-to-end deep learning approach for object affordance detection», in *IEEE International Conference on Robotics and Automation*, (2018) 1–5.

which potentially leads to practical limitations in the long run[23].

Imagination-based AI systems require considerably less training data. The imagination-based chair-detection system described in §1 attained its accuracy of 97% after training on just 30 examples, while the imagination-based container-detection system in Wu and Chirikjian[24] attained the accuracy observed above after training on 11 examples. The imagination-based object-detection system in Hinkle and Olson[25] was able to detect four classes of objects with 80% accuracy after training on 85 examples. Insofar as these examples are representative of the data requirements of imagination-based AI, this approach appears to mitigate a weakness of deep learning.

Another well-known limitation of deep learning is a lack of interpretability. The training of deep learning models usually involves little human intervention—it is left up to the model to identify the relevant features from training examples and relate these features to the target tasks. Consequently, after a model has been trained and deployed, the process by which it arrives at its decisions is often somewhat opaque. While it is possible to track the activation of various nodes in a neural network, it is difficult—often practically impossible—to identify the input features that correspond to the activation of each node. And even if it is in principle possible to implement deep learning models whose working is comprehensible, the more successful deep learning models in practice are usually of such complexity that opacity is practically unavoidable. For this reason, deep learning models are typically labelled 'black boxes', and sometimes thought to have limited utility for the purpose of scientific discovery (Raghu & Schmidt; Rudin)[26]. This lack of explainability has ethical implications for the use of AI systems. For instance, a less explainable system makes it difficult for users to tell when system decisions are incorrect or unreliable, and hence require greater epistemic trust from their users (Buijsman & Veluwenkamp; Gouveia & Malík; Miller)[27]. Besides this, it has also been argued that less explainable AI systems lend themselves less readily to contestation (Henin & Métayer)[28], make algorithmic bias less detectable[29] (Chuan et al.), and subsequently might even infringe on the

[23] N. C. Thompson, , K. Greenewald, K. Lee, G. F. Manso, *The computational limits of deep learning*. arXiv preprint, 2022.

[24] H. Wu, & G. S. Chirikjian, «Can I pour into it? Robot imagining open containability affordance of previously unseen objects via physical simulations», in *IEEE Robotics and Automation Letters, 6*(1), 2020, 271–278.

[25] L.Hinkle, & E. Olson, «Predicting object functionality using physical simulations». *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2013) 2784–2790.

[26] , M. Raghu, & E. Schmidt, «*A survey of deep learning for scientific discovery»*, in arXiv preprint (2020); Rudin, C. «Stop explaining black box machine learning models of high stakes decisions and use interpretable models instead». *Nature Machine Intelligence, 1*(5), (2019) 206–215.

[27] S, Buijsman & H. Veluwenkamp «Spotting When Algorithms Are Wrong» in *Minds and Machines* (2022).**;** S. S. Gouveia & J. Malík «Crossing the trust gap in medical AI: Building an abductive bridge for XAI». *Philosophy and Technology, 37*(3), (2024) 1–25.**;** T. Miller «Explaination in artificial intelligence: Insights from the social sciences», in *Artificial intelligence, 267* (2019) 1–38. Thanks to two reviewers for helping to clarify this point.

[28] C. Henin & D. L. Métayer «Beyond explainability: justifiability and contestability of algorithmic decision systems», in *AI and Society, 37* (2022) 1397–1410.

[29] C. Henin & D. L. Métayer, «Beyond explainability: justifiability and contestability of algorithmic decision systems», in *AI and Society, 37*, (2022) 1397–1410.

autonomy of their users (Bayamlıoğlu; Lyons et al.)[30].

Imagination-based AI systems run into fewer issues with explainability because their inner workings are more perspicuous. Since it is known that such systems frame their tasks in terms of object affordances, the features that influence the systems' decisions are less mysterious—the decisions may be expected to depend primarily on the detection of affordances. And since imagination-based systems detect affordances using internal simulations, we can examine the process by which these systems arrive at their decisions by examining the simulations. When the chair-detection system in §1 classifies an object as a chair, for instance, the simulation results would reveal the orientations of the target object the system considered stable, and the expected behaviour of humans when attempting to sit on the object. This greater interpretability mitigates the limitations of deep learning models with respect to explainability. By having greater insight into the process by which imagination-based systems arrive at their decisions, users are in a better position to tell when the system's results are unreliable and intervene accordingly. This greater explainability is particularly significant in view of the observed performance levels of imagination-based systems above, since there is often thought to be a trade-off between the explainability and performance levels of AI systems (Došilović,, Brčić & Hlupić)[31].

Besides the ability of AI to perform tedious or difficult tasks, we might also be interested in AI as a model of human intelligence. Indeed, it is sometimes thought that one factor contributing to the success of deep learning is the resemblance between the structures of neural networks and human cognition. For instance, the process of visual perception in humans is thought to have a hierarchical structure in which perceived features move from being simple to abstract along the process (Goodale & Milner; Hong et al.)[32]. Similarly, convolutional neural networks process images by passing activation values sequentially through layers of the network, with activation values in nodes corresponding to progressively more complex features in later layers. This resemblance between the two structures is sometimes thought to explain the effectiveness of convolutional neural networks at image recognition tasks (Khaligh-Razavi & Kriegeskorte; Yamins & DiCarlo)[33]. The process by which humans comprehend and summarise an article is also thought to

[30] E. Bayamlıoğlu, «The right to contest automated decisions under the General Data Protection Regulation: Beyond the so-called "right to explanation"», in *Regulation and Governance*. (2021) **;** , H. Lyons, , E.Velloso & , T. Miller «Conceptualising contestability: Perspectives on contesting algorithmic decisions», in *Proceedings of the ACM on Human-Computer Interaction, 5*, (2021) 1–25.

[31] , F. K. Došilović, M. Brčić, & N. Hlupić (2018). «Explainable artificial intelligence: A survey». *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 210–215.**;** Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. (2019). XAI—Explainable artificial intelligence. *Science Robotics, 4*(37) This purported relation has more recently been contested — see for instance Crook et al. (2023)**:** Crook, B., Schlüter, M., & Speith, T. Revisiting the performance-explainability trade-off in explainable artificial intelligence (XAI). *arXiv preprint* (2023) 1–9.

[32] M. A. Goodale & A. D. Milner, «Separate visual pathways for perception and action». *Trends in Neurosciences, 15*(1), (1992) 20–25 **;** , H. Hong, D. L.Yamins, N. J. Majaj & J. J. DiCarlo «Explicit information for category-orthogonal object properties increases along the ventral stream», in *Nature Neuroscience, 19*(4), (2016) 613–622.

[33] S. M. Khaligh-Razavi & N. Kriegeskorte. «Deep supervised, but not unsupervised, models may explain IT cortical representation», in *PLoS Computational Biology, 10*(11) (2014); D. L. Yamins & J. J. DiCarlo, «Using goal-driven deep learning models to understand sensory cortex», in *Nature Neuroscience, 19*(3), (2016) 356–365.

follow a hierarchical structure, which, when implemented as a neural network, yields an abstractive text summarisation model that outperforms alternatives in producing human-like results[34]. From this perspective, the accuracy of deep learning models in reproducing human decisions serves as evidence that neural networks can model human intelligence.

However, there are also reasons to think that deep learning models might be limited as models of human intelligence. For one, as noted above, the number of training examples required to train deep learning models far exceeds the number of examples humans may plausibly be expected to require when learning to perform tasks like object recognition. Also, the low explainability of deep learning models implies that our grounds of claiming resemblance between deep learning models and human intelligence are limited. Even if such models can attain high accuracy levels, this only shows that they can replicate human behaviour at certain tasks; it would be difficult to claim with confidence that the observed behaviour is the result of human-like processes.

In these respects, imagination-based systems better model human intelligence than do deep learning models. As observed earlier, the performance levels of imagination-based systems are comparable to those attained by deep learning models, so the two approaches are on a par with respect to replicating human behaviour. At the same time, the number of training examples required by imagination-based systems to attain these performance levels are typically fewer than a hundred—closer to what human learning may be expected to require. Moreover, their greater explainability allows us to examine the processes by which they produce their behaviour and compare them to human cognition. To be sure, explainability alone does not guarantee that imagination-based AI accurately models human intelligence when it comes to the relevant tasks. Further work would be required to determine if, for instance, humans learn to recognise chairs by considering the stability of objects and the ways they interact with humans.

Nevertheless, there is some reason to think that the processes used by imagination-based systems are human-like in important respects. For one, according to popular views in ecological psychology, animals do learn to interact with their environment by perceiving object affordances. Furthermore, some research in cognitive science suggests that humans sometimes do employ mental processes resembling the internal simulations of imagination-based systems. In an study by Metzler and Shepard[35], for example, human subjects were presented with pairs of similar-looking objects differing by a rotation and asked to determine if the objects are congruent. Subjects' introspective reports suggest that they performed the tasks by mentally rotating the objects, and these reports were corroborated by the observation that subjects' response times to the task increase linearly with the angular displacement of the objects. Other experiments similarly suggest that humans work with mental images

[34] M.Yang, C. Li,, Y. Shen, Q.Wu, Z. Zhao, & X. Chen, Hierarchical Human-Like Deep Neural Networks for Abstractive Text Summarization. *IEEE Transactions on Neural Networks and Learning Systems, 32*(6), (2021) 2744–2757.

[35] J. Metzler, & R.N. Shepard. Mental rotation of three-dimensional objects. *Science, 171*, (1971) 701–703.

when performing certain tasks (Ball et al.; Finke & Pinker)[36] . Some studies moreover suggest that the performance of imagination-based AI is at least partly due to its use of human-like processes. In a study of human shape perception, it was found that the accuracy and speed with which humans infer the shape of an occluded object from visual cues agree with those of AI models based on simulations, even more so than with those of models based on convolutional neural networks (Yildirim et al.)[37]. Lake et al. also argue, along these lines, that aligning AI systems with human intelligence would require that the former move beyond mere pattern-recognition algorithms to the use of causal world models.

The upshot is that the imagination-based approach to AI is reasonably viable, sharing the strengths of deep learning while mitigating some of its weaknesses. To be sure, adopting the imagination-based approach might not require abandoning other approaches. Although the examples of imagination-based AI examined above do not rely at all on neural networks, it might be possible for the imagination-based approach to come in as an enhancement to deep learning. There might be an AI system with the capacity to detect numerous objects via the imagination-based approach, but it might also be too resource-intensive in practice to perform the simulations corresponding to all these possible objects for a single instance of object detection. Such a system could perhaps be made more efficient by having a deep learning model early in the pipeline that narrows down the class of candidate objects before turning to the imagination-based approach to classify among these candidates. If the classification required by the deep learning component of such a system is less extensive than that required of models based purely on deep learning, perhaps the data requirements and complexity of such a system can still be reduced, and the hybrid system would retain some strengths of imagination-based AI.

## 3. Technical limitations

For its strengths, imagination-based AI as it presently stands also has limitations compared to more popular approaches to AI, particularly with respect to generalisability. While deep-learning-based models have found a wide variety of applications, the imagination-based approach has thus far been applied mostly to object recognition tasks. Several obstacles affect the generalisability of the imagination-based approach.

First, the generalisability of imagination-based systems runs into a conceptual challenge, since at present it is difficult to see how some tasks can be made amenable to the imagination-based approach. For instance, it is unclear how text summarisation,

[36] T. Ball, S. Kosslyn & B. Reiser, «Visual images preserve metric spatial information: evidence from studies of image scanning», in *Experimental Psychology: Human Perception and Performance, 4*, (1978) 47–60.; R. A. Finke & S. Pinker «Directional scanning of remembered visual patterns», in *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*(3), (1983) 398–410.

[37] I. Yildirim, M.H. Siegel, A.A. Soltani, S.R. Chaudhari & J.B. Tenenbaum, «3D Shape Perception Integrates Intuitive Physics and Analysis-by-Synthesis». *arXiv preprint* (2023); B.M. Lake, T.D. Ullman, J.B. Tenenbaum & S.J. Gershman, «Building machines that learn and think like people», in *Behavioral and Brain Sciences*, (2017) 1–72.

speech recognition, or facial recognition can be framed in terms of affordances, or even approached using computer simulations. Nevertheless, it is plausible that some form of mental simulation plays a role in these tasks for humans—when attempting to identify a familiar face in an image, humans might recall known faces from memory and mentally compare them with the target image. There is thus the possibility that the applicability of imagination-based AI can be broadened with the right conceptual framing for the target tasks.

Another factor impeding generalisation is the lack of a clear delineation between domain-specific and domain-agnostic components of imagination-based systems. It is known that the deep learning approach allows for *transfer learning*, in which a neural network is pre-trained on a relatively broad dataset spanning a wide domain before it is fine-tuned on a smaller and more domain-specific dataset to yield a final model. For instance, a convolutional neural network can be pre-trained on a large dataset to be able to identify common features present in a wide range of objects, and subsequently fine-tuned on domain-specific images to yield image recognition models for use in specific contexts. An advantage of transfer learning is that it reduces the amount of domain-specific training data required, since the number of training examples required to fine-tune a pre-trained model is typically smaller than that required to train a neural network from scratch. Neural networks allow for training at some level of generalisation because of their hierarchical structure, in which the activation of nodes in early layers of a model typically correspond to simpler features and that of later layers correspond to more domain-specific features. A neural network can thus be pre-trained by having its early layers tuned to the more domain-agnostic features of a wide range of data, and subsequently fine-tuned by training its later layers to recognise domain-specific features.

In contrast, imagination-based systems at present seem tailored particularly to their respective tasks, with task-specific features hardcoded into the system. The chair-recognition system in §1, for instance, has criteria for the stability of the target object and its interaction with humans that seem useful just for chair-detection and not for other tasks. It seems unclear how such a system can be made applicable to other tasks simply by slight modification of its structure, or by tuning the parameters of just some of its components.

Yet another factor hindering the generalisability of imagination-based AI is its present inability to identify the relevant features of data independently. It is known that deep learning can be used for *feature extraction* from a rich dataset with many features, in which a neural network is designed in such a way that when trained on the dataset, the trained model can subsequently be examined to discover what the most representative features of the data are. With images of human faces, for example, one might be interested in identifying the features that play the most significant roles in facial recognition. One type of neural network that can extract the desired features is an *autoencoder*, in which the target output is identical to the input, and the network contains a hidden layer whose dimensionality is much smaller than that of the input layer. After training on the image data, the activation values of the hidden layer for a given image can be taken as an efficient representation of the image. If, moreover, the activation values can be correlated with particular visual features within the images, we would have a sense for which features are the most significant for the model's internal representation. Feature extraction techniques like these have been applied to aid image recognition (Gogoi & Begum; Wang et al.)[38] and natural language processing (Liang et al.)[39]. Deep learning models can perform feature extraction because they work largely without human intervention—during training, it is left to the model to learn what the relevant features of the data are. This opens the possibility that deep learning models might find significance in features that were previously unnoticed by human engineers.

At present, imagination-based systems lack the ability to draw insights from data in this way, since the relevant features for the target tasks are hardcoded in such a way that little room is left for independent identification of those features. For example, it is hardcoded into the chair-detection system in §1 that chairs are characterised by their stability and interaction with humanoids, and learning is limited just to the parameters that define stability and acceptable humanoid interaction. While does not mean that the system cannot produce unexpected results, the heavy reliance on domain-specific information from human engineers makes it doubtful that the system can independently discover new insights from data. Nevertheless, some conceptual work is being done to explore the possibility of reducing the reliance of imagination-based systems on hardcoded domain knowledge (Gan)[40]. If the imagination-based approach can be developed in such a way as to accommodate a greater degree of learning, this would increase its potential to broaden its range of possible applications.

To be sure, the above limitations should not be taken as reasons to disregard the imagination-based approach to AI. Indeed, deep learning has limitations of its own (as observed earlier), and these are often seen not as reasons against the use of deep learning, but as avenues for further research. Similarly, the observations above suggest that imagination-based AI presents several opportunities for technical and conceptual exploration, which can potentially yield significant progress toward making imagination-based AI more broadly applicable.

## 4. Philosophical issues

The similarity between imagination-based AI and human intelligence was earlier identified as a potential strength of the imagination-based approach. To assess this potential strength more thoroughly, some philosophical issues raised by imagination-based AI may be considered.

Foremost, the appropriateness of the term 'imagination' as a description of the internal simulations used by imagination-based systems should be discussed. Wu et

[38] M. Gogoi & S.A. Begum, «Image Classification Using Deep Autoencoders». *IEEE International Conference on Computational Intelligence and Computing Research* (2017).; Y. Wang, X. Wang, & W. Liu, «Unsupervised local deep feature for image recognition», in *Information Sciences, 351*, (2016) 67–75.

[39] H. Liang, X. Sun, Y. Sun & Y. Gao, «Text feature extraction based on deep learning: a review», in *EURASIP Journal on Wireless Communications and Networking, 211* (2017). doi:10.1186/s13638-017-0993-1

[40] N. Gan, «Learning the logic of simulation», in J. Seibt, P. Fazekas, & O. S. Quick (Eds.), *Social Robots With AI: Prospects, Risks, and Responsible Methods,* IOS Press 2025, pp. 428–439.

al. describe their chair-detection system as using «novel method which imagines an object's affordance as a chair by physically simulating a human agent interacting with the object» [41], while Grabner et al. describe a similar system as one that «imagines an actor performing an action typical for the target object class, instead of relying purely on the visual object appearance» [42].

From a philosophical perspective, the resemblance between the relevant simulations and human imagination suggest initial reasons in favour of this term. On some popular accounts, human imagination consists in the running of an internal simulation (Currie & Ravenscroft; Goldman,; Jackson)[43]. Furthermore, imagination is often thought to resemble perception in its phenomenology. Amy Kind, for instance, writes that 'the phenomenology of imagining resembles that of perception' (Kind) [44], while Christopher Gauker takes imaginative episodes to have 'the same representational format as perceptions' (Gauker)[45]. At the same time, imagination is thought to differ from perception by being able to represent things differently from how they in fact are (Jackson)[46]. These key features seem to be shared by the internal simulations of imagination-based AI, which represent actual objects in hypothetical scenarios. The two phenomena thus seem to be analogous at a functional level.

There are also similarities between the epistemic roles of human imagination and the relevant simulations. It has been argued that imaginative episodes can represent future possibilities (Fernández; Harris; Schacter et al.)[47], and our beliefs about such possibilities can be justified by imagination (Jackson; Kind; Kind & Kung; Myers; Williamson) [48].

---

[41] H. Wu, & G.S. Chirikjian, «Can I pour into it? Robot imagining open containability affordance of previously unseen objects via physical simulations», *IEEE Robotics and Automation Letters, 6*(1) (2020) 271.

[42] H. Grabner, J. Gall & L. V. Gool, What makes a chair a chair? *Conference on Computer Vision and Pattern Recognition* (2011) 1529.

[43] G. Currie & I. Ravenscroft, *Recreative Minds: Imagination in Philosophy and Psychology,* Oxford University Press Oxford 2002; A. Goldman, *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading.*: Oxford University Press, Oxford 2006**;** M. B. Jackson, «Justification by imagination», in F. Macpherson & F. Dorsch (Eds.), *Perceptual Imagination and Perceptual Memory,* Oxford University Press, Oxford 2018, pp. 209–226.

[44] (Kind, 2001, p. 96)

[45] C. Gauker. «Imagination constrained, imagination constructed», in *Inquiry* (2021) 2**.** Also see G. Currie, & I. Ravenscroft. *Recreative Minds: Imagination in Philosophy and Psychology,* Oxford University Press, Oxford 2002; P. Fazekas, B. Nanay & J. Pearson «Offline perception: an introduction». *Philosophical Transactions B,* (2021) *376*; M. B. Jackson, «Justification by imagination», in F. Macpherson & F. Dorsch (Eds.), *Perceptual Imagination and Perceptual Memory*, Oxford University Press, Oxford 2018, pp. 209–226; F. Macpherson & F. Dorsch, *Perceptual Imagination and Perceptual Memory,* Oxford University Press, Oxford 2016, p. 2.

[46] M. B. Jackson Justification by imagination. In F. Macpherson & F. Dorsch (Eds.), *Perceptual Imagination and Perceptual Memory* , Oxford University Press, Oxford 2018, pp. 211–214.

[47] J. Fernández, «The contents of imagination», *Canadian Journal of Philosophy, 52*(8), (2023) 828–842.; P. L. Harris, (2021), «Early constraints on the imagination: The realism of young children», in *Child Development, 92*(2) 466–483; D.L. Schacter, Addis & R. L. Buckner, R. L., «Remembering the past to imagine the future: the prospective brain», *Nature Reviews Neuroscience, 8*, (2023) 657–661.

[48] M.B. Jackson, Justification by imagination. In F. Macpherson & F. Dorsch (Eds.), *Perceptual Imagination and Perceptual Memory,* Oxford University Press, Oxford 2018, pp. 209–226**; A.** Kind, How imagination gives rise to knowledge. In F. Macpherson & F. Dorsch (Eds.), *Perceptual Memory and Perceptual Imagination,* Oxford University Press, Oxford, 2018, pp. 227–246 **; A.** Kind, & P. Kung, *Knowledge Through Imagination,* Oxford University Press, Oxford 2016 ; J. Meyers, Reasoning with imagination. In A. Kind & C. Badura (Eds.), *Epistemic Uses of Imagination*: Routledge, London 2021; **T.** Williamson, Knowing by imagining. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination,* Oxford University Press, Oxford 2016. For arguments

Likewise, the simulations run by imagination-based AI systems are intended to be informative about future possibilities in our world—when the chair-detection system runs simulations in which a humanoid attempts to sit on an object, these are intended to represent what would happen if a human were to attempt to sit on the actual object.

Moreover, the AI simulations and human imagination are both reliably informative under similar conditions. It has been argued that human imagination can justify beliefs about our world only insofar as they align with reality in certain key respects. On Amy Kind's account of 'imagining under constraints '[49], for instance, imaginative episodes are justificatory only insofar as they satisfy two constraints: a reality constraint requiring the content of imaginative episodes to correspond to the world, and a change constraint requiring imaginative episodes to develop in a realistic manner. On alternative accounts, similarly, imagination justifies belief only insofar as it is constrained to align with reality (Dorsch; Jackson; Langland-Hassan; Meyers; Miyazono & Tooming; Williams; Williamson)[50]. Likewise, the simulations run by imagination-based AI systems can reliably inform the systems' decisions only insofar as they are sufficiently accurate in their representation of the target scenario, the physical laws of our world, and the like.

To be sure, the resemblance between AI simulations and human imagination is limited. *Prima facie*, the former operate at a purely mechanical and syntactic level, while the latter is naturally associated with intentionality and semantic understanding, which are not typically associated with AI systems. Furthermore, since imagination is a mental state, it might be thought that imagination requires consciousness. But it would be at least surprising if imagination-based AI systems are thought to be conscious simply because they use internal simulations. Similarly, we also use human imagination to take the perspectives of others (Ravenscroft)[51], to engage in fiction (Currie; Walton)[52]**,** and to engage emotionally with scenarios (Leeuwen)[53], all of which capacities are not usually attributed to AI systems. These implications suggest that when AI simulations

---

to the contrary, see J. Egeland, «Imagination cannot justify empirical belief», *Episteme, 18*(4), (2021) 507–513. and Spaulding, S. (2016). Imagination through knowledge. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination* (pp. 207–226). Oxford: Oxford University Press).

[49] A**.** Kind', «Imagining under constraints», in A. Kind & P. Kung (Eds.), *Knowledge Through Imagination*, Oxford University Press, Oxford 2016, pp. 145–159, A. Kind, «How imagination gives rise to knowledge», in F. Macpherson & F. Dorsch (Eds.), *Perceptual Memory and Perceptual Imagination,* Oxford University Press, Oxford 2018, pp. 227–246.

[50] F. Dorsch, «Knowledge by imagination—how imaginative experience can ground factual knowledge», *Teorema: Revista Internacional de Filosofía, 35*(3), (2016) 87–116.; M. B. Jackson, «Justification by imagination», In F. Macpherson & F. Dorsch (Eds.), *Perceptual Imagination and Perceptual Memory,* Oxford University Press, Oxford 2018, pp. 209–226; P. Langland-Hassan, «On choosing what to imagine». In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination, op. cit., v*ol. 2, Oxford University Press, Oxford 2016, pp. 61–84 ; J. Meyers Reasoning with imagination. In A. Kind & C. Badura (Eds.), *Epistemic Uses of Imagination*: Routledge, London 2021; K. Miyazono & U. Tooming, «Imagination as a generative source of justification», *Noûs* (2023); D. Williams, «Imaginative constraints and generative models», *Australasian Journal of Philosophy, 99*(1), (2021) 68–82; T. Williamson, «Knowing by imagining», in A. Kind & P. Kung (Eds.), *Knowledge Through Imagination,* Oxford University Press, Oxford 2016.

[51] J. Ravenscroft, «What is it like to be someone else? Simulation and empathy». *Ratio, 11*, (2002) 170–185.

[52] G. Currie, *The Nature of Fiction*. Cambridge University Press, Cambridge 1990 ; K. Walton, *Mimesis as Make-Believe*. Harvard University Press, Massachusetts (1990).

[53] N. v. Leeuwen, «The imaginative agent», in A. Kind & P. Kung (Eds.), *Knowledge Through Imagination,* op.cit., pp. 85–110.

are described as instances of imagination, either the description should be interpreted as a mere metaphor indicating only an analogy at the functional level, or these cases give us reason to reevaluate our usual understanding of the capabilities of AI systems.

The upshot is that although imagination-based AI systems are so-called in the engineering literature, investigation is required to determine if this term is appropriate. And whether we admit imagination among AI systems might have implications for further issues regarding imagination or the capabilities of AI. For instance, in the debate over the nature of mental imagery, some hold that mental images consist primarily in propositional representations (Pylyshin)[54], while others take them to consist primarily in pictorial representations (Kosslyn)[55] , and yet others hold intermediate views (Tye)[56]. If AI simulations are considered instances of imagination, disputants in the imagery debate might have to account for the possibility of AI imagination in formulating their view. In particular, proponents of the pictorial view would have to explain how, while simulations are being run by an AI system, the data stored in the system's memory can be a picture-like representation.

So apart from its technical aspects, imagination-based AI also presents possible areas for philosophical exploration, with potential payoffs for other issues in the philosophy of AI or mind.

## 5. Conclusion

This paper aimed to highlight the imagination-based approach as a recent and relatively unknown approach to AI. Although presently less popular, imagination-based AI shares the strengths of deep learning when it comes to performance, while mitigating issues related to data requirements and explainability. And as models of human intelligence, there are reasons to think that imagination-based systems more closely resemble human intelligence than do deep learning models. The imagination-based approach thus presents a viable alternative (or supplement) to deep learning in at least some cases.

But work remains to be done to develop the imagination-based approach to AI. At present, imagination-based AI has limited applicability, and some conceptual and technical work is required to overcome this limitation and allow imagination-based AI to attain the same breadth of application as deep learning. As imagination-based systems attain human-level performance at a greater range of tasks, they might bear in new ways on familiar issues regarding the nature of imagination and the capabilities of AI systems. As an area for research, therefore, imagination-based AI provides several avenues for technical and philosophical exploration.

[54] Z. Pylyshin, «The imagery debate: Analogue media versus tacit knowledge», *Psychological Review,* 88(1), (1981) 16–45.

[55] S. M. Kosslyn, «Scanning Visual Images: Some Structural Implications», *Perception and Psychophysics,* 14, (1973). 90–94.

[56] M. Tye, *The Imagery Debate*. Massachusetts, MIT Press (1991).

## Bibliographical References


Ball, T., Kosslyn, S., & Reiser, B. (1978). Visual images preserve metric spatial information: evidence from studies of image scanning. *Experimental Psychology: Human Perception and Performance, 4*, 47–60.

Bar-Aviv, E., & Rivlin, E. (2016). Functional 3D Object Classification Using Simulation of Embodied Agent. *Proceedings of the British Machine Vision Conference*, 307–316.

Bayamlıoğlu, E. (2021). The right to contest automated decisions under the General Data Protection Regulation: Beyond the so-called "right to explanation". *Regulation and Governance*.

Buijsman, S., & Veluwenkamp, H. (2022). Spotting When Algorithms Are Wrong. *Minds and Machines*. doi:10.1007/s11023-022-09591-0.

Chong, I., & Proctor, R. W. (2019). On the evolution of a radical concept: Affordances according to Gibson and their subsequent use and development. *Perspectives on Psychological Science, 15*(1), 117–132.

Chuan, C., Sun, R., Tian, S., & Tsai, W. S. (2024). EXplainable Artificial Intelligence (XAI) for facilitating recognition of algorithmic bias: An experiment from imposed users' perspectives. *Telematics and Informatics, 91*, 1–15.

Crook, B., Schlüter, M., & Speith, T. (2023). Revisiting the performance-explainability trade-off in explainable artificial intelligence (XAI). *arXiv preprint*, 1–9.

Currie, G. (1990). *The Nature of Fiction*. Cambridge: Cambridge University Press.

Currie, G., & Ravenscroft, I. (2002). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford: Oxford University Press.

Do, T., Nguyen, A., & Reid, I. (2018). Affordancenet: An end-to-end deep learning approach for object affordance detection. *IEEE International Conference on Robotics and Automation*, 1–5.

Dorsch, F. (2016). Knowledge by imagination—how imaginative experience can ground factual knowledge. *Teorema: Revista Internacional de Filosofía, 35*(3), 87–116.

Došilović, F. K., Brčić, M., & Hlupić, N. (2018). Explainable artificial intelligence: A survey. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 210–215.

Egeland, J. (2021). Imagination cannot justify empirical belief. *Episteme, 18*(4), 507–513.

Erdemir, E., Frankel, C. B., Kawamura, K., Gordon, S. M., Thornton, S., & Ulutas, B. (2008). Towards a cognitive robot that uses internal rehearsal to learn affordance relations. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016–2021.

Fazekas, P., Nanay, B., & Pearson, J. (2021). Offline perception: an introduction. *Philosophical Transactions B, 376*. doi:10.1098/rstb.2019.0686

Fernández, J. (2023). The contents of imagination. *Canadian Journal of Philosophy, 52*(8), 828–842.

Finke, R. A., & Pinker, S. (1982). Spontaneous imagery scanning in mental extrapolation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8*(2), 142–147.

Finke, R. A., & Pinker, S. (1983). Directional scanning of remembered visual patterns. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*(3), 398–410.

Gan, N. (2025). Learning the logic of simulation. In J. Seibt, P. Fazekas, & O. S. Quick (Eds.), *Social Robots With AI: Prospects, Risks, and Responsible Methods* (pp. 428–439): IOS Press.

Gauker, C. (2021). Imagination constrained, imagination constructed. *Inquiry*. doi:10.1080/00

20174X.2021.1933748.
Gibson, J. J. (1950). *The Perception of the Visual World*. Boston: Houghton Mifflin.
— (1979). *The Ecological Approach to Visual Perception*. New York: Routledge.
Gogoi, M., & Begum, S. A. (2017). Image Classification Using Deep Autoencoders. *IEEE International Conference on Computational Intelligence and Computing Research*.
Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.
Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15*(1), 20–25. doi:10.1016/0166-2236(92)90344-8.
Gouveia, S. S., & Malík, J. (2024). Crossing the trust gap in medical AI: Building an abductive bridge for XAI. *Philosophy and Technology, 37*(3), 1–25.
Grabner, H., Gall, J., & Gool, L. V. (2011). What makes a chair a chair? *Conference on Computer Vision and Pattern Recognition*, 1529–1536. doi:10.1109/CVPR.2011.5995327.
Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. (2019). XAI—Explainable artificial intelligence. *Science Robotics, 4*(37).
Gupta, S. (2016). *Deep learning vs. traditional machine learning algorithms used in credit card fraud detection.* (Master thesis). National College of Ireland.
Ha, D., & Schmidhuber, J. (2018). World models. *arXiv preprint*, 1–21. doi:10.5281/zenodo.1207631.
Harris, P. L. (2021). Early constraints on the imagination: The realism of young children. *Child Development, 92*(2), 466–483. doi:10.1111/cdev.13487.
Henin, C., & Métayer, D. L. (2022). Beyond explainability: justifiability and contestability of algorithmic decision systems. *AI and Society, 37*, 1397–1410.
Hinkle, L., & Olson, E. (2013). Predicting object functionality using physical simulations. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2784–2790.
Ho, S. B. (2023). Why is that a good or not a good frying pan? Knowledge representation for functions of objects and tools for design understanding, improvement, and generation. *arXiv*. doi:10.48550/arXiv.2303.06152.
Hong, H., Yamins, D. L., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience, 19*(4), 613–622. doi:10.1038/nn.4247.
Jackson, M. B. (2018). Justification by imagination. In F. Macpherson & F. Dorsch (Eds.), *Perceptual Imagination and Perceptual Memory* (pp. 209–226). Oxford: Oxford University Press.
Kanezaki, A., Matsushita, Y., & Nishida, Y. (2018). Rotationnet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5010–5019.
Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology, 10*(11).
Kind, A. (2001). Putting the image back in imagination. *Philosophy and Phenomenological Research, 62*(1), 85–109.
— (2016). Imagining under constraints. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination* (pp. 145–159). Oxford: Oxford University Press.
— (2018). How imagination gives rise to knowledge. In F. Macpherson & F. Dorsch (Eds.), *Perceptual Memory and Perceptual Imagination* (pp. 227–246). Oxford: Oxford University Press.

Kind, A., & Kung, P. (2016). *Knowledge Through Imagination*. Oxford: Oxford University Press.

Kosslyn, S. M. (1973). Scanning Visual Images: Some Structural Implications. *Perception and Psychophysics, 14*, 90–94.

Lai, Y. (2019). A Comparison of Traditional Machine Learning and Deep Learning in Image Recognition. *Journal of Physics: Conference Series, 1314*, 1–8.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 1–72.

Langland-Hassan, P. (2016). On choosing what to imagine. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination* (Vol. 2, pp. 61–84). Oxford: Oxford University Press.

Leeuwen, N. v. (2016). The imaginative agent. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination* (pp. 85–110). Oxford: Oxford University Press.

Li, H. (2018). Deep learning for natural language processing: advantages and challenges. *National Science Review, 5*(1), 24–26.

Liang, H., Sun, X., Sun, Y., & Gao, Y. (2017). Text feature extraction based on deep learning: a review. *EURASIP Journal on Wireless Communications and Networking, 211*. doi:10.1186/s13638-017-0993-1.

Lyons, H., Velloso, E., & Miller, T. (2021). Conceptualising contestability: Perspectives on contesting algorithmic decisions. *Proceedings of the ACM on Human-Computer Interaction, 5*, 1–25.

Macpherson, F., & Dorsch, F. (2016). *Perceptual Imagination and Perceptual Memory*. Oxford: Oxford University Press.

Meng, X., Wu, H., Ruan, S., & Chirikjian, G. S. (2023). Prepare the chair for the bear! Robot imagination of sitting affordance to reorient previously unseen chairs. *arXiv preprint*. doi:arXiv:2306.11448.

Metzler, J., & Shepard, R. N. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.

Meyers, J. (2021). Reasoning with imagination. In A. Kind & C. Badura (Eds.), *Epistemic Uses of Imagination*: Routledge.

Miller, T. (2019). Explaination in artificial intelligence: Insights from the social sciences. *Artificial intelligence, 267*, 1–38.

Miyazono, K., & Tooming, U. (2023). Imagination as a generative source of justification. *Noûs*. doi:10.1111/nous.12458

Myers, J. (2021). The epistemic status of the imagination. *Philosophical Studies, 178*, 3251–3270.

— (forthcoming). How imagination informs. *Philosophical Quarterly*.

Paterakis, N. G., Mocanu, E., Gibescu, M., Stappers, B., & Alst, W. v. (2017). Deep Learning Versus Traditional Machine Learning Methods for Aggregated Energy Demand Prediction. *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe*, 1–6.

Pylyshin, Z. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88*(1), 16–45.

Raghu, M., & Schmidt, E. (2020). *A survey of deep learning for scientific discovery*. arXiv preprint.

Ravenscroft, I. (2002). What is it like to be someone else? Simulation and empathy. *Ratio, 11*, 170–185.

Rudin, C. (2019). Stop explaining black box machine learning models of high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience, 8*, 657–661.

Schmidhuber, J. (2015). On learning to think: Algorithmic information theory for novel combinations of reinforcement learning controllers and recurrent neural world models. *arXiv preprint*, 1–36. doi:10.48550/arXiv.1511.09249.

Seita, D., Florence, P., Tompson, J., Coumans, E., Sindhwani, V., Goldberg, K., & Zeng, A. (2021). Learning to Rearrange Deformable Cables, Fabrics, and Bags with Goal-Conditioned Transporter Networks. *IEEE International Conference on Robotics and Automation*, 4568–4575.

Spaulding, S. (2016). Imagination through knowledge. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination* (pp. 207–226). Oxford: Oxford University Press.

Su, J. C., Gadelha, M., Wang, R., & Maji, S. (2018). A deeper look at 3d shape classifiers. *Second Workshop on 3D Reconstruction Meets Semantics*.

Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2022). *The computational limits of deep learning*. arXiv preprint.

Tye, M. (1991). *The Imagery Debate*. Massachusetts: MIT Press.

Walton, K. (1990). *Mimesis as Make-Believe*. Massachusetts: Harvard University Press.

Wang, Y., Wang, X., & Liu, W. (2016). Unsupervised local deep feature for image recognition. *Information Sciences, 351*, 67–75.

Williams, D. (2021). Imaginative constraints and generative models. *Australasian Journal of Philosophy, 99*(1), 68–82.

Williamson, T. (2016). Knowing by imagining. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination*. Oxford: Oxford University Press.

Wu, H., & Chirikjian, G. S. (2020). Can I pour into it? Robot imagining open containability affordance of previously unseen objects via physical simulations. *IEEE Robotics and Automation Letters, 6*(1), 271–278.

Wu, H., Meng, X., Ruan, S., & Chirikjian, G. S. (2022a). *Put the bear on the chair! Intelligent robot interaction with previously unseen chairs via robot imagination*. arXiv preprint.

Wu, H., Misra, D., & Chirikjian, G. S. (2020). Is that a chair? Imagining affordances using simulations of an articulated human body. *IEEE International Conference on Robotics and Automation*.

Wu, H., Ye, J., Meng, X., Paxton, C., & Chirikjian, G. S. (2022b). *Transporters with visual foresight for solving unseen rearrangement tasks*. arXiv preprints.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., & Xiao, J. (2015). 3d shapenets: A deep representation for volumetric shapes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1912–1920.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience, 19*(3), 356–365. doi:10.1038/nn.4244.

Yang, M., Li, C., Shen, Y., Wu, Q., Zhao, Z., & Chen, X. (2021). Hierarchical Human-Like Deep Neural Networks for Abstractive Text Summarization. *IEEE Transactions on Neural Networks and Learning Systems, 32*(6), 2744–2757. doi:10.1109/TNNLS.2020.3008037.

Yildirim, I., Siegel, M. H., Soltani, A. A., Chaudhari, S. R., & Tenenbaum, J. B. (2023). 3D Shape Perception Integrates Intuitive Physics and Analysis-by-Synthesis. *arXiv preprint*. doi:arXiv:2301.03711.

Zhu, Y., Zhao, Y., & Zhu, S. (2015). Understanding Tools: Task-Oriented Object Modeling, Learning and Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 2855–2864. doi:10.1109/CVPR.2015.7298903.