LAF | https://doi.org/10.15302/J-LAF-2026-0057
Volume 14 / Issue 6 / 260057 / 2026
Received Date | 2026-07-21 Accepted Date | 2026-09-16
© Higher Education Press 2026

Beyond Interpolation and Extrapolation: The Place of Imagination in the Age of Artificial Intelligence —A Discussion of Liyan Xu's *Interpolative and Extrapolative Work—The Place of Human Intelligence in the Age of Artificial Intelligence*

Hongmou ZHANG[1,2,*]

1 School of Government, Peking University, Beijing 100871, China
2 Institute of Urban Governance, Peking University, Beijing 100871, China

*CORRESPONDING AUTHOR
Address: No. 5 Yiheyuan Road, Haidian District, Beijing 100871, China
Email: zhanghongmou@pku.edu.cn

ABSTRACT

Liyan Xu proposes that human intellectual work can be divided into interpolative and extrapolative work, holds that artificial intelligence (AI) is, at its core, an interpolation engine, and argues that interpolative work should be delegated to AI. This article proposes that extrapolation also contains another kind of activity, namely problem reframing. It is characterized by proposing new directions in the absence of existing references. Most such attempts do not hold up, while a few change the state of the relevant field. This article calls this capacity imagination and advances two arguments. First, the value of imagination arises outside existing evaluation standards, and imagination is the part of extrapolation that is hardest for AI to absorb through data accumulation and relationship fitting. Second, imagination is not a capacity that humans stably possess. After AI enters everyday work, whether the conditions for practicing it contract or are redistributed depends on the order of the workflow, and this article holds that sustaining this capacity requires corresponding social arrangements. The article discusses, in turn, the sources and costs of imagination, the contraction and redistribution of the conditions for practicing imagination under AI use, and the value judgments contained in the division between interpolation and extrapolation. It concludes that the human capacity for exploration that departs from existing standards, together with the corresponding institutional arrangements, should be protected and cultivated.

KEYWORDS

Imagination; Problem Reframing; Interpolation and Extrapolation; Human–AI Division of Labor; Artificial Intelligence; Value Judgment

HIGHLIGHTS

- Problem reframing within extrapolation is least readily absorbed by AI as data accumulate
- Imagination entails value-based selection and commitment; generation alone is insufficient to constitute imagination
- AI's training of human expression may constrain the practice of imagination, depending on workflow order
- Interpolation also contains judgment, and delegating it to AI transfers some aesthetic trade-offs
- Human capacity to challenge norms and explore the unknown should be cultivated, with supporting workflows and institutions

1 Introduction

In *Interpolative and Extrapolative Work—The Place of Human Intelligence in the Age of Artificial Intelligence* (hereafter *Interpolative and Extrapolative Work*), Liyan Xu proposes that human intellectual work falls into two kinds, interpolative and extrapolative. The former operates within the boundaries of what is already known and fills in details, and the latter poses new questions and forges connections that did not previously exist. Current artificial intelligence (AI) is, at its core, an interpolation engine, so interpolative work should be delegated to AI, while human effort concentrates on the extrapolative frontier[1]. This article broadly accepts this judgment, but argues that extrapolation contains a kind of activity that merits separate discussion, namely the reframing

of problems. Problem reframing is still, in essence, a form of extrapolation. The word "beyond" in the title refers to the fact that the nature of this activity exceeds a presupposition implicit in the dichotomy between interpolation and extrapolation, namely that extrapolative work "will" gradually turn into interpolative work as data accumulate. In fact, within extrapolative work, problem reframing is precisely the part that is hardest to convert in this way.

Sometimes the posing of new questions and the forging of new connections between bodies of knowledge take place between two known fields, and once such a new connection is identified, it can be quickly summarized, learned, and replicated. At other times, however, raising a new question is closer to probing in the dark without any reference. One first proposes a direction that seems far-fetched. Most attempts end in failure, but a very few arrive at a position that no one had foreseen, and thereby change the state of an entire research field. This article calls the latter capacity imagination and, on this basis, puts forward two arguments. First, imagination is the part of extrapolation that is hardest for AI to absorb, because its value arises precisely outside existing evaluation standards. Second, imagination is not an innate gift that humans stably possess. After AI enters everyday work, the conditions under which people practice imagination may contract or be redistributed, depending on how workflows are arranged. Sustaining human imagination is therefore a value orientation that a society must actively adopt. The following sections attempt to argue for these two points.

2 Problem Reframing Within Extrapolation

One of the most direct manifestations of imagination is the reframing of the problem itself (sometimes what happens first is its deconstruction). Such reframing is not limited to major scientific revolutions, and instances of it can also be found in everyday life. For example, there is an online youth community whose interest centers on homophonic puns and linguistic deconstruction[2], and it represents a certain "abstract culture" among contemporary youth[3]. A common move in this culture is to re-segment a familiar word. An English counterpart is to read "important" as "import" plus "ant," that is, an ant that has been imported.① Such a reading unexpectedly does not extend outward along the original semantic relations, and instead changes the category to which the word belongs (cognitive linguistics calls this process frame-shifting[4]). Repeating a ready-made pun is mere imitation, and playing to the crowd does not in itself constitute innovation. The value of this kind of "doing abstract" behavior lies in the moment of spark when someone first notices that "important" can be read as an imported ant, and that spark in itself carries a deep meaning of innovation. Viktor Shklovsky called this practice of interrupting habitual perception defamiliarization (остранение)[5]. Wolfgang Pauli used the phrase "not even wrong" to describe propositions so absurdly wrong that they were not worth discussing in terms of right or wrong②[6]. Following Pauli, a creative answer that deconstructs the question may be called "not even right" (in the sense of going beyond merely being right). The respondent temporarily steps outside the correctness standard of the original question and turns to examining whether that standard itself holds.

Such questioning of existing definitions also appears in spatial practice. Every year, some temporary settlements are built in a desert during a specific period and dismantled after the event ends. These settlements have roads, blocks, public services, and operating rules, and accommodate tens of thousands of temporary residents for a short period. If such a settlement still cannot be considered a city, what other conditions must a city meet? Is permanence an indispensable condition of a city? This example raises the question of what a city is[7]. Problem reframing in landscape architecture and spatial design often begins with a reexamination of established premises: reimagining the possibilities of an existing site and even going beyond the existing site to explore the possible forms of human settlement. Such questions rarely appear in the design briefs issued by a client, yet they may inspire discussions of the forms of human existence, and concern for the fate of humanity that matter more than the project itself.

Such problem reframing has already been discussed in the literature. Donald Schön called the process by which practitioners redefine the boundaries of a task "problem setting," and pointed out that the situations practitioners face must often first be given a frame for action before they become problems that can be solved[8]. Horst Rittel and Melvin Webber's account of "wicked problems" goes further. Planning problems often have no uncontested definition, and every formulation of a problem already contains a value position[9]. Margaret Boden distinguished three types of creativity (combinational, exploratory, and transformational), of which transformational creativity changes the very rule space that generates solutions[10]. These three bodies of work, from different

① The example in the Chinese version of this article is "明知山有虎，那就不去明知山" ("There are tigers on Mingzhi Mountain, so don't go to Mingzhi Mountain"). It plays on the saying "明知山有虎，偏向虎山行" ("knowing full well that there are tigers in the mountains, one still heads for them"), in which "明知" means "knowing full well." The joke, which has circulated widely in this community, re-segments "明知山" and reads it as the name of a mountain.

② A well-known example of "not even wrong" is answering "an obtuse angle" to the question "What is 1 + 1?"

disciplines, point to the same conclusion. Problem reframing is an independent intellectual operation, and it cannot be reduced to searching for better answers within a given problem (or generalizing outward within a certain range).

This conclusion bears directly on the dichotomy drawn in *Interpolative and Extrapolative Work*. Taking renderings as an example, Xu points out that extrapolative work that once occupied the frontier is gradually absorbed into the interpolable as data accumulate. This process holds for the connection-forging type of extrapolation distinguished at the beginning of this article. Once such connections are stably established, they enter the range of what AI can learn. Extrapolation of the problem-reframing type is different. It challenges the evaluation standards themselves, and data accumulated under the old standards cannot be used to understand a rejection of those standards. Problem reframing is therefore the hardest to bring into the scope of interpolation as data accumulate.

It should be noted that problem reframing is present throughout human work but is very sparse. Most human research and design also proceed within existing frameworks, and a large number of attempts at departure end in failure. Everyday language calls a successful moment of departure a "spark," and the word itself contains a judgment about frequency. Sparse does not mean secondary. The directions of knowledge and design are often reset precisely by a few successful departures. Yet sparsity also has a methodological consequence for AI. Large language models are built on data volume, and the more frequently an event occurs, the more easily it is recognized as a pattern. Sparks of problem reframing sit in the low-frequency tail of any corpus, and each successful reframing changes the distribution of subsequent data, so earlier samples can hardly be used to predict the next one. Methods aimed at finding regularities have difficulty characterizing such events, and the evaluation of imagination cannot rely on average (or maximum likelihood) metrics. Its effect on the overall direction of human knowledge is concentrated in a few events that cannot be represented by probability. The natural next question, then, is where this capacity to depart from existing standards comes from.

3 Sources and Costs of Imagination

On the sources of imagination, the history of philosophy has offered some relatively conservative definitions. David Hume held that the materials of thought come from sensation and experience, and that all the mind can do is compound, transpose, augment, or diminish them[11]. The performance of generative AI shows precisely why this definition is incomplete. The combinatorial capacity of contemporary large language models already far exceeds that of any individual. If imagination consisted only in combining old materials, models with massive combinatorial capacity should already possess imagination. Hume's definition therefore needs a supplement, and the supplement lies in selection based on aesthetics (or value judgment). Discussing mathematical creation, Henri Poincaré pointed out that the number of possible combinations is enormous, and that the real work lies in how the few "useful and beautiful" connections come to be selected, a selection that depends on the mathematician's aesthetic sensibility[12]. Current models use statistical likelihood as the selection criterion in generation, and the aesthetic selection Poincaré described is not within their objective function. The sparsity described in Section 2 has a direct consequence here. Events that are low-frequency and that change the distribution sit exactly where likelihood-based selection has the greatest difficulty reaching (what geometry calls a cusp). Charles Peirce's notion of abduction further clarifies the nature of this selection. Faced with a surprising fact, the researcher first proposes an explanation that has not yet been verified, and invests work in it before the evidence arrives[13]. A large model can generate any number of candidate explanations on instruction, but provisional trust in, and sustained investment in, one of these explanations cannot be supplied by the act of generation itself (although this can of course be done under human guidance). Taken together, imagination reconfigures the relations among materials and makes a committed selection among the results of that reconfiguration, and the materials themselves need not be new.

On this basis, imagination can be broken down into three stages, namely generating candidates that depart from existing standards, making a value-based selection among the candidates, and sustaining investment in the absence of support from existing evaluation standards. On instruction, AI can also generate candidates that are low-probability, unconventional, and even marked by frame-shifting. The first stage is therefore not unique to humans, and the difference between imagination as understood in this article and the generative capacity of AI is concentrated in the latter two stages, that is, selection and commitment. The example given in Section 2 can also be understood in this way. The reading of "important" as an imported ant could perhaps be enumerated by a model, but noticing that this reading is worth keeping, and that it carries a distinctive value in a social sense, is a human choice.

Such selection has observable behavioral features. In a longitudinal study of art students, Jacob Getzels and Mihaly Csikszentmihalyi found that students who explored their subject matter more openly and for longer in the early phase were later

more likely to produce works judged to be original[14]. High-level creators (note that the notion of "high-level" itself implies a value judgment) tend to settle the final form of the problem later. The counterpart in design education is that site investigations and sketches in the concept phase often go through several rounds of iteration before the designer settles on how to formulate the design problem, and the concept phase is precisely the part that design review criteria have the most difficulty covering. This feature runs opposite to the current training direction of large models. Large models are optimized to give complete and acceptable answers as quickly as possible, and deferring the settlement of the problem has no corresponding place in their training objectives. Simulated annealing can provide a limited analogy for the difference between these two ways of answering. The algorithm accepts temporarily worse moves (even random moves) with a certain probability, so that the search does not remain stuck in a local optimum[15]. This is close to the off-the-point answers discussed above, in that both first leave the currently highest-rated position and enter a region that for the time being does not make sense. The analogy is limited because simulated annealing still obeys a given objective function, whereas human imagination may bring the objective function itself under doubt. The algorithm moves only within a given solution space, while humans can question the grounds on which that solution space is set.③

Questioning the solution space itself leads imagination toward its mature form, and also toward its costs. René Descartes's method of doubt requires that all doubtful beliefs be temporarily suspended until an indubitable starting point is found[16]. Willard Van Orman Quine further proposed that knowledge faces the test of experience as a whole, and that any part of it is in principle revisable[17]. Karl Popper placed falsifiability at the core of the status of scientific theories[18]. The three hold different positions, but all refuse to treat existing answers as the final boundary, and those who undertake such refusal must also bear the corresponding costs. Research may fail, a site design may not be accepted, and both the designer's professional reputation and the public interest may suffer. Existing AI can list counterexamples on instruction, but it has no reason to take on this risk of its own accord. Its generation of a sentence of objection differs in nature from researchers giving up a theory they have long believed. It can thus be seen that imagination depends on a set of fragile conditions. People must be willing to remain in uncertainty, must keep room for asking bad questions and drawing failed sketches, and must also be prepared to pay the price of failure.

This article cannot yet offer a complete explanation of the "ultimate" source of imagination. In terms of the three stages distinguished above, as to why a spark appears at a given moment, existing accounts can mostly offer only post hoc narratives. Discussing the Schema, Immanuel Kant pointed out that the Schema is always a product of the imagination (Einbildungskraft), and that the schematism of the understanding is a hidden art in the depths of the human soul (eine verborgene Kunst in den Tiefen der menschlichen Seele), whose true operations are hard to reveal[19]. Popper likewise pointed out that the initial conception of a new idea does not fall within the scope of logical analysis, and that there is no logical method of arriving at new ideas[18]. Section 2 described problem reframing as probing in the dark precisely because its starting point is hard to reduce to a function of past accumulation. What can be explained are the latter two stages. Judging whether a departure is worth persisting in and sustaining investment in depends on the practitioner's familiarity with the boundaries of existing frameworks, and this familiarity comes from the experience of working within those frameworks firsthand. Schön pointed out that a practitioner's capacity for problem setting rests on a "repertoire" accumulated across a large number of concrete cases. Facing a new situation, the practitioner sees it as a variant of some familiar case and thereby redefines the task[8]. The aesthetic selection described by Poincaré likewise rests on long mathematical training[12]. Only by revising a design again and again as it repeatedly fails to hold up can a practitioner tell which rules are real constraints and which are merely habits of the profession. What interpolative work supports, therefore, are the two stages of selection and commitment. The spark itself is not produced by accumulation, but without accumulation, people have no way of telling which spark is worth investing in. This is consistent with the claim in Section 2. Problem reframing is hard to bring into the scope of interpolation as data accumulate precisely because its starting point lies outside accumulation. What remains to be shown is that, as AI enters everyday work, this set of conditions is changing.

4 Contraction and Redistribution of Imagination

The change comes first from a training process that runs in the opposite direction. Humans train AI, and AI also trains humans. Current assistant-type AI models are trained with human feedback and tend to give answers that are relevant, helpful, clear, safe, and consistent with instructions[20]. These requirements improve the

③ In theory, large models can also learn and train the hyperparameters on which they depend, that is, "lift" the problem up by one level. What this article seeks to point out, however, is the human capacity for "unlimited lifting."

usability of the tools, and they also form a dominant model of expression, with complete sentences, balanced structure, and an even-handed attitude. As users keep reading, accepting, and revising such outputs, over time they easily come to treat the model's "average" expression as normal expression, and to take high-probability AI lines of argument as the way thinking "has always been."

There is preliminary evidence for this concern. Anil Doshi and Oliver Hauser found that generative AI can raise the average ratings of individual short stories while reducing the overall diversity of a set of works[21]. A cross-cultural experiment by Dhruv Agarwal et al. showed that AI writing suggestions make the writing of different participants more similar, and push the expression of Indian participants toward the American writing style with which the model is more familiar[22]. Another experiment on poetry found that when people were placed in the position of revising finished AI output, their creativity was lower than in independent creation, and the gap disappeared only when the system placed people in the position of co-creators[23]. These studies are not yet sufficient to prove that AI use necessarily harms imagination, but they already suffice to show that an improvement in the quality of works and an improvement in creative capacity are not necessarily the same thing.

In design disciplines, the same convergence may appear earlier, in problem setting. Highly reflective water surfaces, lush vegetation, and clean paving may make a single AI-generated image attractive enough, but with extensive use, such images gradually prescribe what kind of landscape is "worth" imagining. If designers start from generated images and then add site-based justification, they easily work backward from answers familiar to the model to the problem, and judgments that should have been formed through investigation, drawing, and repeated back-and-forth with the site are compressed into the screening of ready-made images. In this article we call this change "the interpolativization of humans." People increasingly occupy the supplementary positions reserved in model outputs, supplementing paragraphs, selecting schemes, and polishing images, while taking less and less part in deciding why a problem is posed the way it is. Section 2 pointed out that problem reframing is inherently sparse in intellectual activity. The other half of the reason can be added here, which is that people's opportunities to practice it are also shrinking. Telling which departures are worth persisting in depends on the experience of working within frameworks firsthand. When such work is handed over to AI as a whole, what diminishes is not only the opportunity to practice departure but also the conditions for forming this discernment.

The crux of the matter, therefore, is not whether design students or paper authors use AI, but the order of the workflow, that is, whether people call on tools after forming a problem, or start from the problems and answers the tools provide. Protecting imagination does not require rejecting AI, but it does require that people keep the parts of the workflow that cannot be outsourced. These include engaging with the site in person, posing the initial question themselves, giving explanations when values conflict, and keeping failed schemes. Once it is acknowledged that the way tools are used in turn shapes capacities, the division of labor between humans and AI can no longer be drawn solely on the basis of technical performance. This also means that the practice conditions discussed in Section 3 need not only contract after AI enters work. Humans have conditions that AI lacks. They perceive the physical world in person, experience spatial relationships on site, and form judgments through bodily experience, and these conditions do not disappear because AI enters the workflow. Once interpolative work such as drawing, retrieval, and text editing is taken over by AI, designers can in principle gain more time to stay on site and explore openly, that is, the conditions for longer exploration of problems in the sense of Getzels and Csikszentmihalyi[14]. The two effects of contraction and release coexist, and AI use itself does not determine which prevails. What determines it is where the released time goes, and whether the workflow places people before or after the formation of the problem. The judgment of this article is therefore conditional. In workflows that start from answers provided by tools, the practice conditions contract, and the evidence of homogenization cited above describes precisely this kind of workflow. In workflows where people first form the problem and then call on tools, the practice conditions may be redistributed, or even expanded. Accordingly, the normative claim of this article is not limited to protecting a capacity that may be lost, and also includes cultivating this capacity through workflow and educational arrangements. In design education, site investigation, open-ended sketching, and the archiving of "failed" schemes all belong to such arrangements. This judgment should also distinguish the stage people are at. For practitioners who have already formed discernment, the time released by handing over repetitive work can be directed toward the workflow in which the problem is formed first. For students still in the learning stage, the same handover directly reduces accumulation itself, and this is precisely why design education needs to keep the parts in which basic work is done by hand.

5 Value Judgments Behind the Division of Labor

The proposal to delegate interpolative work to AI presupposes that interpolation is work in which judgment has already been

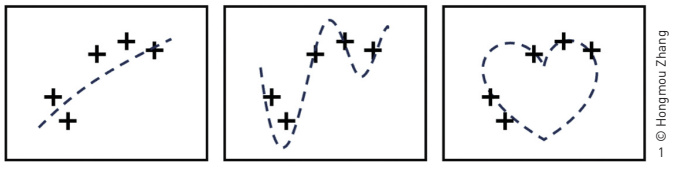
made and only execution remains. This presupposition also deserves scrutiny. In model fitting, given a number of known points, one can draw many (even infinitely many) curves that pass through them (Fig. 1), and the known points do not tell us how the points "should" be connected. The concept of inductive bias in machine learning shows that a model must adopt some selection tendency in advance, beyond the data, or it cannot settle on one among the many functions that fit the data equally well[24]. The no-free-lunch theorem shows that, detached from specific problems, there is no algorithm that is best for all optimization tasks[25]. In design practice, under the same set of site conditions, there are still many schemes that satisfy technical codes, and whether to prioritize flood control or habitat is itself a ranking of values. Interpolation, then, also contains judgment, and delegating interpolation to AI as a whole also hands over part of the aesthetic judgments and trade-offs that ought to be made by humans.

Following this line of reasoning, determining whether a piece of work has already become interpolative is itself a value judgment, and the questions of who sets the criteria for this determination and on what basis they are adopted are easily passed over in discussion. Leaving technology to evolve on its own and letting the market choose appears to take no position, but in fact hands value judgments over to those who control models, data, computing power, and platforms. Langdon Winner pointed out long ago that technical artifacts are not neutral instruments detached from politics and institutions, and that they entrench particular power relations and modes of action[26]. Further, for society, imagination is not only an individual quality but also a capacity at the organizational level of governments, firms, communities, and other types of organizations, and the conditions for imagination available to different parties are unequal. When the development of AI is led by those who control models, data, and computing power, this inequality is itself part of the value judgments behind the division of labor. Whose imagination is protected and who defines the futures worth imagining, on the one hand, and which work should be handed over to AI, on the other hand, are two sides of the same question. The workflows and institutions discussed in Section 4 are therefore not only arrangements of individual work habits, but also a distribution of the conditions for imagination among different social actors.

AI is still changing rapidly, and the boundary between interpolation and extrapolation is also moving. The composition of training data, the choice of objective functions and evaluation standards, and the direction of investment and regulation all influence where this boundary moves. Research on responsible innovation calls for anticipating impacts in advance during research and development, listening to the views of different groups, and adjusting direction in response to feedback[27]. The notion of "sociotechnical imaginaries" in science and technology studies also reminds us that technological trajectories always carry visions of a shared future[28]. Value guidance therefore cannot consist only of aligning with the past. Historical data preserve knowledge, but they also preserve biases and institutional inertia, and fitting the past more accurately is sometimes merely replicating the past more effectively. Iason Gabriel's analysis of value alignment distinguishes among people's immediate instructions, actual preferences, real interests, and the values they endorse upon reflection[29]. The mainstream values of a society are not the average of existing preferences, and should also include criticism of preferences themselves, as well as the spirit of an era (Zeitgeist).

Finally, we return to the two arguments put forward at the beginning of this article. Imagination is the part of extrapolation that is hardest to absorb as data accumulate, and it is also the part most dependent on practice conditions, and whether these conditions contract or are redistributed after AI enters everyday work depends on where the workflow places people. *Interpolative and Extrapolative Work* provides a valuable starting point for discussing the division of labor between humans and AI. What this article seeks to add is that what truly needs to be protected and cultivated in this division of labor is not any specific type of job or skill, but the human capacity to depart from existing standards and explore in the dark, together with the workflows and institutions that protect and cultivate this capacity. How these workflows and institutions are distributed among different actors is likewise part of the question of the division of labor. If, amid convenience, people gradually lose the patience to remain in uncertainty, and are no longer willing to ask questions that may be wrong, or to draw a sketch that will most likely fail, then before being replaced by AI, people will already have trained themselves into more competent interpolators. Which work AI should take on is a question that must ultimately be answered by humans, and for which humans should be responsible.

Fig. 1 Different ways of fitting the same set of data.



© Hongmou Zhang

Competing interests | The author declares that he has no competing interests.

REFERENCES

[1] Xu, L. (2026). Interpolative and extrapolative work—The place of human intelligence in the age of artificial intelligence. *Landscape Architecture Frontiers*, *14*(4), 260032.

[2] He, R., He, Y., Bai, L., Liu, J., Sun, Z., Tang, Z., … Deng, N. (2024). Chumor 1.0: A truly funny and challenging Chinese humor understanding dataset from Ruo Zhi Ba. *arXiv*. arXiv:2406.12754.

[3] Zhu, S., Zhu, Z., & Whyke, T. W. (2025). From live-streaming to memes: An exploratory study of the Chinese abstract culture as a structure of feeling among Chinese youth. *Journal of Youth Studies*, 1–17.

[4] Coulson, S. (2001). *Semantic Leaps: Frame-Shifting and Conceptual Blending in Meaning Construction*. Cambridge University Press.

[5] Shklovsky, V. (1965). Art as Technique. In: L. T. Lemon & M. J. Reis (Eds.), *Russian Formalist Criticism: Four Essays* (pp. 3–24). University of Nebraska Press.

[6] Peierls, R. E. (1960). Wolfgang Ernst Pauli, 1900–1958. *Biographical Memoirs of Fellows of the Royal Society*, *5*(1), 174–192.

[7] White, C. L. (2020). *The Archaeology of Burning Man: The Rise and Fall of Black Rock City*. University of New Mexico Press.

[8] Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. Basic Books.

[9] Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*, 155–169.

[10] Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms* (2nd ed.). Routledge.

[11] Hume, D. (1748/2000). *An Enquiry Concerning Human Understanding*. Oxford University Press.

[12] Poincaré, H. (1910). Mathematical creation. *The Monist*, *20*(3), 321–335.

[13] Peirce, C. S. (1878). Deduction, induction, and hypothesis. *Popular Science Monthly*, *13*, 470–482.

[14] Getzels, J. W., & Csikszentmihalyi, M. (1976). *The Creative Vision: A Longitudinal Study of Problem Finding in Art*. Wiley.

[15] Kirkpatrick, S., Gelatt, C. D., Jr., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680.

[16] Descartes, R. (1641/2017). *Meditations on First Philosophy: With Selections from the Objections and Replies* (J. Cottingham, Ed. & Trans., 2nd ed.). Cambridge University Press.

[17] Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review*, *60*(1), 20–43.

[18] Popper, K. R. (1959/2002). *The Logic of Scientific Discovery*. Routledge.

[19] Kant, I. (1781/1998). *Critique of Pure Reason* (P. Guyer & A. W. Wood, Eds. & Trans.). Cambridge University Press.

[20] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., … Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, *35*, 27730–27744.

[21] Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, *10*(28), eadn5290.

[22] Agarwal, D., Naaman, M., & Vashistha, A. (2025). AI suggestions homogenize writing toward Western styles and diminish cultural nuances. In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1117.

[23] McGuire, J., De Cremer, D., & Van de Cruys, T. (2024). Establishing the importance of co-creation and self-efficacy in creative collaboration with artificial intelligence. *Scientific Reports*, *14*, 18525.

[24] Mitchell, T. M. (1980). *The Need for Biases in Learning Generalizations*. Rutgers University, Department of Computer Science, CBM-TR-117.

[25] Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82.

[26] Winner, L. (1980). Do artifacts have politics?. *Daedalus*, *109*(1), 121–136.

[27] Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580.

[28] Jasanoff, S., & Kim, S.-H. (Eds.). (2015). *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. The University of Chicago Press.

[29] Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines*, *30*, 411–437.

LAF | https://doi.org/10.15302/J-LAF-2026-0057
第14卷 / 第6期 / 260057 / 2026
收稿日期 | 2026-07-21　录用日期 | 2026-09-16
中图分类号 | TU984　文献标识码 | B

内插与外推之外：AI时代想象力的位置
——与许立言《内插性工作与外推性工作——AI时代的人类智识定位》一文商榷

张洪谋[1,2,*]

1 北京大学政府管理学院，北京 100871
2 北京大学城市治理研究院，北京 100871

***通信作者**
地址：北京市海淀区颐和园路5号
邮编：100871
邮箱：zhanghongmou@pku.edu.cn

摘要

许立言提出，人类智识活动可分为内插与外推，认为人工智能（AI）本质上是内插引擎，内插性工作应由AI处理。本文提出，外推中还包含另一类活动，即问题重构，其特征是在缺少既有参照时提出新方向，多数尝试未能成立，少数会改变相关领域的态势。本文将这种能力称为“想象力”，并提出两个论点。第一，想象力的价值产生于既有评价标准之外，是外推中最难通过数据积累与关系拟合被AI吸收的部分。第二，想象力并非人类稳定具备的能力，AI进入日常工作后其练习条件是收缩还是重新分配，取决于工作流程的顺序，本文认为维护这种能力需要相应的社会安排。本文依次讨论想象力的来源与代价，AI使用与想象力练习条件的收缩与再分配，以及内插与外推的划分中包含的价值判断，最后指出应保护并培养人类偏离既有标准的探索能力与相应的制度安排。

关键词

想象力；问题重构；内插与外推；人机分工；人工智能；价值判断

文章亮点

- 外推中的问题重构最难随数据积累被AI吸收
- 想象力是基于价值判断的选择与承诺，生成本身不足以构成想象力
- AI训练人类表达习惯的过程可能压缩想象力的练习条件，其影响取决于工作流程的顺序
- 内插同样包含判断，委托AI意味着转移部分审美取舍
- 应培养人类偏离既有标准、探索未知的能力，并建立相应的流程与制度

1 引言

许立言在《内插性工作与外推性工作——AI时代的人类智识定位》（以下简称《内插与外推》）一文中提出，人类的智识活动可分为内插与外推两类。前者在既有知识边界内补全细节，后者提出新问题、建立此前不存在的联系，而现阶段的人工智能（AI）本质上是一个内插引擎，因此应当把内插性工作交给AI，人类则专注于外推的前沿[1]。本文大体接受这个判断，但认为外推之中还包含一种值得单独讨论的活动，即对问题的重构。问题重构本质上仍属于外推。标题中的“之外”指的是，这类活动的性质超出了内插与外推二分所隐含的一个预设，即外推性工作“会”随数据积累逐步转化为内插性工作。但实际上，在外推性工作中，问题重构恰好是最难完成这一转化的部分。

有些时候提出新问题、建立新的知识联系是在两个已知领域之间发生的，这类新联系一旦被识别，即可被快速总结、学习和复制。但另一些时候，新问题的提出更接近在没有参照的黑暗中试探，先给出一个看似不着边际的方向，多数尝试以失败告终，但极少数会到达一个此前无人预见的位置，并因此改变整个研究领域的态势。本文把后一种能力称为“想象力”，并由此提出两个观点。其一，想象力是外推之中最难被AI吸收的部分，因为它的价值恰恰产生于既有评价标准之外。其二，想象力并非人类稳定拥有的天赋，AI进入日常工作之后，人类练习想象力的条件面临收缩，也可能被重新分配，这取决于工作流程的安排，维护人类想象力因而是社会必须主动确立的价值导向。以下试论证这两个观点。

2 外推之中的问题重构

想象力最直观的表现形式之一，是对问题本身的重构（有时首先发生的是解构）。这种重构不仅限于重大的科学变革，日常生活中也存在一定实例。例如，某个以谐音双关和语言解构为兴趣导向的青年网络社群[2]，代表了当代青年的某种“抽象文化”[3]。该文化中有一句近来流传较广的双关解构：“明知山有虎，那就不去明知山”，把成语里的“明知山”重新断句，读成了一个地名。这个句子出人意料地未沿着原有语义关系向外延伸，而更换了词语所属的范畴（认知语言学将这个过程称为“框架转换”[4]）。重复现成的双关只是模仿，哗众取宠本身也不构成创新，但这类“搞抽象”行为的价值在于——有人第一次察觉到可以把“明知山”作为一个地名使用的灵机一动时刻，这种灵机一动本身即具有深刻的创新内涵。维克托·什克洛夫斯基把这种打断习惯感知的做法称为“陌生化”（остранение）[5]。沃尔夫冈·泡利曾用“not even wrong”（“甚至不能称之为错误”）评价一些错到离谱，以至于不值得讨论对错的命题①[6]。参考泡利的说法，充满创造力的问题解构式回答可以称为“not even right”（可对照译为“甚至不仅仅是正确”），回答者暂时退出了原问题的正确性标准，转而检查这个标准本身是否成立。

这种对既有定义的质疑也出现在空间实践中。例如，一些临时聚落会在每年的特定时期在沙漠中集中建成，并在活动结束后拆除。这类聚落可能具备道路、街区、公共服务设施和运行规则，并在短期内容纳大量临时居民。如果一个具备这些特征的聚落不能被视为城市，那么城市还需要满足哪些条件？长期存在是否是城市不可缺少的条件？该临时聚落的案例促使我们重新思考“城市是什么”的问题[7]。景观设计与空间设计中的问题重构，往往始于对既定前提的重新审视：重新想象既有场地的可能性，甚至超越既有场地，对人类聚落的可能形态进行讨论。这类问题通常不会出现在委托方给定的规划设计项目任务书中，却可能启发比项目本身更重要的、关于人类生存形态的讨论和对于人类命运的关切。

这类问题重构行为在既有文献中已有讨论。唐纳德·舍恩将实践者重新界定任务边界的过程称为“问题设定”，并指出实践者面对的情境往往需要先被赋予一个行动框架，然后才能成为可以求解的问题[8]。霍斯特·里特尔和梅尔文·韦伯关于“抗解问题”（wicked problem）的论述则进一步指出，规划问题往往不存在无争议的定义，对问题的每一次表述本身已经包含价值立场[9]。玛格丽特·博登则把创造力分为组合性、探索性和转换性3类，其中转换性创造力改变的正是生成方案的规则空间本身[10]。尽管来自不同学科，这3项研究均指向同一结论：问题重构是一种独立的智识操作，无法被还原为在既定问题内部（或在一定范围内向外泛化后）寻找更优答案的过程。

① “not even wrong”的典型例子：当被问及“1+1 等于几”时，回答“钝角”。

这个结论直接关系到《内插与外推》一文的二分方式。许立言以效果图绘制为例，指出曾经位于前沿的外推性工作会随数据积累逐渐被纳入内插的范畴。这个过程对本文开头区分的建立联系类的外推是成立的，这类联系一旦稳定建立，即进入可被AI学习的范围。问题重构类外推则不同，它挑战的是评价标准本身，而旧标准下积累的数据，无法用来理解一次对旧标准的拒绝。因此，问题重构最难随数据积累而被纳入内插的范围。

需要说明的是，问题重构在人类工作中普遍存在但非常稀疏，多数人类研究与设计同样在既有框架内推进，大量偏离尝试也以失败告终。日常语言中把成功的偏离时刻称为灵机一动，这个说法本身就包含频率上的判断。稀疏并不等于次要，知识与设计的方向往往正是由少数成功的偏离重新设定的。然而，稀疏却同时带来一个AI方法论上的后果。大语言模型以数据量为基础，事件出现得越频繁，越容易被识别为模式，而问题重构类的灵机一动在任何语料中都处于低频的尾部，并且每一次成功的重构都会改变此后数据的分布，因此先前的样本难以用于预测下一次。以找规律为目标的方法难以刻画这类事件，对想象力的评估也无法依赖平均指标，它对人类知识整体走向的作用集中在少数无法由概率代表的事件上。那么，自然而然的下一个问题是：这种偏离既有标准的能力从何而来？

3 想象力的来源与代价

关于想象力的来源，哲学史提供过一些较为保守的界定。大卫·休谟指出，思想的材料来自感觉和经验，心灵能做的只是对它们进行组合、移置、增加和减少[11]。生成式AI的表现恰好说明这个界定为何不完整。当代大语言模型的组合能力已经远超任何个人，如果想象力仅在于组合旧材料，具备海量组合能力的模型就应当已经具备想象力。因此休谟的界定需要补充，补充的位置在于基于审美（或称为价值判断）的选择。亨利·庞加莱讨论数学创造时指出，可能的组合数量巨大，真正的工作在于少数“有用而美”的联系何以被选中，而这种选择依赖数学家的审美感受[12]。当前模型以统计似然作为生成时的选择标准，庞加莱所说的审美选择并不在其目标函数之内，而第2节所述的稀疏在这里有直接的后果，低频且改变分布的事件，恰好处于似然选择最难覆盖的位置（几何中所谓“尖点”）。查尔斯·皮尔士所说的溯因进一步说明了这种选择的性质，研究者在面对意外事实时先提出一个尚未被证实的解释，并在证据到来之前为它投入工作[13]。大模型可以在指令下生成任意数量的候选解释，但对其中某一个解释的暂时信任与持续投入，无法由生成行为本身提供（当然可以在人类引导下进行）。综合来看，想象力重构的是材料之间的关系，并对重构的结果作出有承诺的选择，材料本身不必是新的。

据此可以把想象力拆分为3个环节——生成偏离既有标准的候选，从

候选中作出价值判断上的选择，以及在缺乏既有评价标准支持的情况下持续投入。AI在指令下同样可以生成低概率、反常规甚至带有框架转换特征的候选，因此第一个环节并非人类独有，本文所说的想象力与AI生成能力的差别集中在后两个环节，即选择与承诺。第2节所举的例子也可以按此理解，“明知山”作为地名的读法可能可以由模型枚举出来，但察觉到这个读法值得保留并具有社会意义下的独特价值，是人类的选择。

这种选择有其可观察的行为特征。雅各布·格策尔斯和米哈伊·奇克森特米哈伊对艺术学生的长期研究发现，前期对研究对象进行更开放、更长时间探索的学生，后来更可能产生被评价为原创的作品[14]，高水平（注意“高水平”这个概念本身就隐含价值判断）的创作者往往更晚才确认问题的最终形态。设计教育中的对应现象是，方案构思阶段的场地调研与草图往往经过多轮迭代之后，设计者才确定设计问题的表述，而方案构思阶段恰好是评图标准最难覆盖的部分。这一特征与当前大模型的训练方向相反，大模型被优化为尽快给出完整、可接受的回答，推迟确认问题在其训练目标中没有对应的位置。模拟退火算法（simulated annealing）可以为这两种回答方式的差别提供一个有限的类比：该算法以一定概率接受暂时更差的移动（甚至是随机移动），以避免搜索停留在局部最优[15]，这与前文所举的答非所问案例接近，都是先离开当前评价最高的位置，进入暂时无法说通的区域。这个类比的有限性在于，模拟退火依然服从既定的目标函数，人的想象力却可能把目标函数本身纳入怀疑。算法只在既定的解空间内移动，人则可以质疑该解空间本身的设定依据②。

对解空间本身的质疑，把想象力引向它的成熟形态，也引向它的代价。勒内·笛卡尔的方法怀疑要求把一切可疑的信念暂时悬置，直到找到不可怀疑的起点[16]。威拉德·范·奥尔曼·奎因进一步提出，知识作为整体面对经验的检验，其中任何部分原则上都可修正[17]。卡尔·波普尔则把可证伪性置于科学理论地位的核心[18]。三人立场各异，但都拒绝把既有答案当作最终边界，而承担这种拒绝的人，也要承担相应的代价——研究可能失败，场地设计方案可能不被接受，设计师的职业声誉与公共利益都可能受损。现有AI可以按指令列举反例，却没有理由主动承担这种风险，它生成一句反对意见，与研究者放弃自己长期相信的理论，性质并不相同。由此可以看出，想象力依赖一组并不牢固的条件，人要愿意停留在不确定中，要保留提出坏问题和画失败草图的空间，也要准备好为失败付出代价。

关于想象力的“终极”来源，本文尚无法提供完整的解释。按前文区分的3个环节来看，一次灵机一动为什么在此时出现，现有的说法大多只能给出事后的叙述。伊曼努尔·康德在讨论图式（Schema）时指出，图式始终是想象力（Einbildungskraft）的产物，而知性的图式法是隐藏在人类心灵深处的一种技艺，其真实的运作方式很难被揭示[19]。波普尔同样指出，一个新想法的最初构想不属于逻辑分析的范围，不存在获得新想法的逻辑方法[18]。上文第2节把问题重构描述为在黑暗中的试探，正是因为它的起点难以还原为既往积累的函数。可以说明的是后两个环节。判断一次偏离是否值得坚持并为之持续投入，依赖从业者对既有框架边界的熟悉，而这种熟悉来自亲自在框架内工作的经历。舍恩指出，从业者的问题设定能力依托于其在大量具体案例中积累的“经验储备”（repertoire），面对新情境时，从业者把它看作某个已知案例的变体，进而重新界定任务[8]。庞加莱所描述的审美选择同样建立在长期的数学训练之上[12]。反复修改一个方案而屡次不成立，从业者才能分辨哪一条规则是真的约束，哪一条只是行业习惯。因此，内插性工作所支持的是选择与承诺这两个环节，灵机一动本身不由积累产生，但缺少积累，人无从分辨哪一次灵机一动值得投入。这与第2节的论断一致，问题重构之所以难以随数据积累被纳入内插的范围，正在于它的起点不在积累之内。接下来要说明的是，AI进入日常工作之后，这组条件正在发生变化。

4 想象力的收缩与再分配

变化首先来自一个方向相反的训练过程。人训练AI，AI也在训练人。当前的助手型AI模型经过人类反馈训练，倾向于给出相关、有用、清楚、安全并符合指令的回答[20]，这些要求提高了工具的可用性，也形成了一种占主导地位的表达范例，句子完整，结构均衡，态度周全。使用者不断阅读、接受和修改这类输出，久而久之容易把模型的“平均”表达当成正常表达，把高概率的AI论证逻辑当成“从来如此”的思考。

这种担忧已有初步证据。阿尼尔·多希和奥利弗·豪瑟发现，生成式AI可以提高个人短篇故事的平均评价，却降低一组作品的整体差异[21]。德鲁夫·阿格拉瓦尔等的跨文化实验显示，AI写作建议会使不同参与者的文字更加相似，并推动印度参与者的表达靠近模型更熟悉的美国写作方式[22]。另一项诗歌实验发现，当人被放在修改AI成品的位置时，其创造性表现低于独立创作。只有当系统把人置于共同创作者的位置，这一差距才会消失[23]。这些研究尚不足以证明使用AI必然损害想象力，但已经足以说明，作品质量的提升与创造能力的提升，未必是一回事。

对于设计学科，同样的趋同可能更早出现在问题设定上。高反射水面、繁茂植被、洁净铺装，单张AI生成图或许足够漂亮，大量使用之后却会逐渐规定什么样的景观才“值得”被想象。设计者如果从生成图开始，再补充场地论证，就容易从模型熟悉的答案倒推问题，原本应由调查、绘图和反复踏勘场地形成的判断，被压缩为对现成图像的筛选。本文把这种变化称为“人的内插化”，人逐渐承担起模型输出中预留的补充位置——补充段落、挑选方案、润色图像，却越来越少参与决定问题

② 理论上大模型也可以对模型依赖的超参数进行学习和训练，即把问题“上提”一层。但本文想指出的是人类“无限上提”的能力。

为何这样提出。第2节曾指出，问题重构在智识活动中本就稀疏，此处可以补充其另一层原因：人们练习问题重构的机会也在减少。分辨哪一次偏离值得坚持，依赖亲自在框架内工作的经历，当这些工作整体转交AI，减少的就不只是练习偏离的机会，还有形成这种分辨力的条件。

因此，问题的关键并非设计学学生或论文作者是否使用AI，而在于工作流程的顺序：人是在形成问题之后调用工具，还是从工具提供的问题与答案开始。保护想象力不要求拒绝AI，但要求人在流程中保留无法外包的环节——亲自接触场地，自己提出最初的问题，在价值冲突时给出解释，保留失败的方案。而一旦承认使用方式会反过来塑造能力，人机分工就无法只依据技术性能来划定。这也意味着，第3节所说的练习条件在AI进入工作之后未必只会收缩。人具备AI不具备的条件，包括亲身感知物理世界、在场地中体会空间关系、通过身体经验形成判断等，这些条件并不因AI进入工作流程而消失。当绘图、检索、文本整理等内插性工作被AI承接后，设计者原则上可以获得更多驻留场地、开放探索的时间，也就是格策尔斯和奇克森特米哈伊所说的具备更长时间探索问题的条件[14]。收缩与释放两种效应同时存在，AI使用本身并不决定哪一种占优，决定因素是被释放的时间流向何处，以及流程把人放在形成问题之前还是之后。因此，本文的判断是条件性的。在从工具提供的答案开始的流程中，练习条件收缩，前文所引的同质化证据描述的正是这种流程。在人先形成问题、再调用工具的流程中，练习条件可能得到重新分配，甚至扩张。相应地，本文的规范性主张也不限于保护一种可能流失的能力，还包括通过流程与教育安排培养这种能力。设计教育中场地调研、开放式草图与“失败”方案的存档，都属于这类安排。这一判断还应区分人所处的阶段。对已经形成分辨力的从业者，转交重复劳动所释放的时间可以投向前一类流程。对尚在学习阶段的学生，同样的转交直接减少了积累本身，因此设计教育需要保留亲手完成基础工作的环节。

5 分工背后的价值判断

把内插性工作交给AI，这个建议预设了内插是判断已经完成、只剩执行的工作。这个预设也值得检视。拟合模型中，给定若干已知点，可以画出很多条（甚至无数条）经过这些点的曲线（图1），已知点并不会告诉我们点与点之间“应当”怎样连接。机器学习中的“归纳偏置”（inductive bias）概念说明，模型必须在数据之外预先采用某种选择倾向，否则无法从众多同样符合数据的函数中确定一个[24]。“没有免费午餐”定理（no-free-lunch theorem）则说明，脱离具体问题，不存在对一切优化任务都最好的算法[25]。落到设计实践中，同一组场地条件之下满足技术规范的方案依然很多，优先防洪还是优先保护生境，本身是价值排序。可见，内插内部同样包含判断。把内插整体交给AI，也意味着把本该由人完成的审美判断与取舍一并交了出去。

沿着这个思路推论，判定一项工作是否已经内插化，本身也是一种价值判断，而判断标准由谁设定、依据什么被采纳，很容易在讨论中被略过。放任技术自然演化、让市场自行选择，看起来没有立场，实际上是把价值判断交给了掌握模型、数据、算力和平台的一方。兰登·温纳早已指出，技术物并非脱离政治与制度的中性器具，它们会固化特定的权力关系与行动方式[26]。进一步，对于社会来说，想象力不仅是个体素养，同样也是政府、企业、社群等组织层面的能力，各方拥有的想象条件并不均等。当AI的发展由掌握模型、数据与算力的一方主导时，这种差异本身就是分工背后价值判断的一部分。谁的想象力得到保护，由谁来界定值得想象的未来，与哪些工作应当交给AI是同一个问题的两面。因此，第4节所说的流程与制度，不只是个体工作习惯的安排，也是不同社会主体之间想象条件的分配。

AI仍在快速变化，内插与外推之间的边界也在移动。训练数据的构成、目标函数与评价标准的选择、投资与监管的方向，都在影响这条边界往哪里移动。“负责任创新”的研究要求研发过程提前考虑影响、听取不同群体意见并根据反馈调整方向[27]。技术社会学所说的“社会技术想象”也提示我们，技术路线始终携带着对共同未来的设想[28]。因此，价值引导不能只做对齐过去的工作，历史数据保存了知识，也保存了偏见与制度惯性。更准确地拟合过去，有时只是更有效地复制过去。伊阿宋·加布里埃尔对价值对齐的分析区分了人的即时指令、现实偏好、实际利益和经过反思后认可的价值[29]。社会的主流价值不等于现有偏好的平均数，还应包含对偏好本身的批评，以及一个时代的时代精神（Zeitgeist）。

最后，回到本文开头提出的两个观点。想象力是外推之中最难随数据积累而被吸收的部分，同时也是最依赖练习条件的部分，而这些条件在AI进入日常工作之后是收缩还是重新分配，取决于工作流程把人放在什么位置。《内插与外推》为讨论人机分工提供了一个有价值的起点，本文想补充的是，这场分工中真正需要保护和培养的，并非某类具体的岗位或技能，而是人类偏离既有标准、在黑暗中探索的能力，以及保护并培养这种能力的流程与制度。这些流程与制度在不同主体之间如何分配，同样是分工问题的一部分。人如果在便利中逐渐失去停留在不确定中的耐心，不再愿意提出可能错误的问题，不再愿意画一张多半会失败的草图，那么在被AI取代之前，人已经先把自己训练成了更合格的内插者。AI应当承担哪些工作，这个问题最终要由人来回答，也应当由人来负责。

图 1. 对于同样一组数据的不同拟合方式。