

一百个点子不等于一个创新：随机生成的五道上限与六道工序

一个下午能生成一百个点子，而创新的产量没变。因为点子从来不是瓶颈——点子是原料。从原料到一件能上战场的东西，中间有六道工序，而机器把其中一道压到了零，另外五道一动没动。

王德生 著 · 学术创造专栏 · 创造力 · 约 6300 字

摘要

生成式人工智能把创意的产量提高了两个数量级，而企业的创新产出没有相应变化。主流解释是执行跟不上、组织不敏捷、创新文化不足。本文提出一个更简单也更能堪的诊断：点子从来不是创新的瓶颈——点子是原料，不是成品，而这一点在点子昂贵的年代被彻底掩盖了：那时一个人一个季度只想得出三个点子，于是“想出点子”看起来就是创新的核心动作。本文给出随机生成的五道上限（局部性：它不知道自己在问题空间的什么位置；无生长力判别：它给你一百个，不告诉你哪个能长；不自动改姓：它给的东西带着来源的口音，进不了目标战场；无长程一致性：明天再问它给你另一百个，跟今天的不接；不生成问题空间：它只在你给的问题里回答，不会告诉你这个问题问错了），并对照六道工序（抓核、抓裂缝、重组、改姓、锻造、投放）指出机器压缩的只是第三道。文章给出一个算术：六道工序的时间比例约为 3:4:1:2:5:1，机器把其中那个“1”压到近乎零，总时间从 16 降到 15——提速约百分之六。而所有人以为它提速了一百倍。文章据此论证“更好的模型救不了它”（五道上限里有三道在原理上不属于生成），给出可压缩与不可压缩时间的分界（沉浸、发酵、打磨三样不可压，而它们恰好占了十六份中的十二份），以及复合创新体的正确分工。文末列出四个失效条件，其中最要紧的一条：六道工序这个划分本身可能是一张画错的地图。

关键词 随机创新上限；六爪动作链；创新工序；复合创新体；时间经济学；SDE

一家公司的创新工作坊，一个下午。以前的产出是：十二个人，四小时，白板上贴了三十几张便利贴，其中大概三个值得再想想。

现在的产出是：一个人，二十分钟，一百个点子。每一个都成句，有逻辑，有初步论证，写得比那三十几张便利贴好。

两年过去了。这家公司推出的新东西，和两年前一样多。

这不是个别现象。它足够普遍，普遍到已经有了一套标准解释：执行跟不上、组织不够敏捷、创新文化不足、缺乏容错机制。

这些解释都在说同一件事：点子有了，是后面掉链子了。

本文要说的是另一件事，而且它更简单：

点子从来不是瓶颈。点子是原料。

>

而这一点在点子昂贵的年代被彻底掩盖了：那时一个人一个季度只想得出三个点子，于是"想出点子"看起来就是创新的核心动作。

一、原料与成品

先立一个区分。

一个点子是什么？它是一个"如果……会怎样"。它可以很聪明，可以很新颖，可以让人眼前一亮。

一个创新是什么？它是一个可调用的新实体——可激发、可释放、可复现。别人能拿它去做事；它在现实里被碰过并且没碎；它有一个能被指着说"就是这个"的形态。

这两者之间隔着的不是"执行"。执行是把一个已经成形的东西做出来。而一个点子还没有成形——它连自己是什么都还说不清。

隔着的是工序。

在点子昂贵的年代，工序是不可见的，因为它和生成是长在一起的：你想一个点子的时候，你已经在判它了；你之所以只想出三个，正是因为另外二十个在你脑子里成形的一瞬间就被你自己毙了。那个自毙的动作、那个"这个方向能长"的直觉、那个"这话得换个说法才有人听"的调整——全部藏在"想点子"这三个字里面，从来没有被单独看见过。

机器把生成剥离出来了。于是那些一直藏在里面的工序，第一次以裸的形式露了出来——而所有人都以为它们不存在。

二、五道上限

随机生成有五道上限。它们不是"当前模型的局限"，其中至少三道在原理上不属于生成这个动作。

上限一·局部性。

它给你一个点子，它不知道这个点子在整个问题空间的什么位置。它不知道这个方向已经有三十家公司试过并且死了；它也不知道这个方向恰好是你们三年前踩过的那个坑。它给的是一个局部的东西，而它对全局一无所知——不是因为它没读过，是因为"位置"不是文本里的信息，位置是相对于你的问题空间而言的，而它没有你的问题空间。

上限二·无生长力判别。

它给你一百个。哪个能长？

它不知道。而更麻烦的是：它给出的一百个在表面质量上是均匀的——每一个都成句、有逻辑、有论证。均匀恰恰是最坏的情况：如果一百个里有九十个明显是垃圾，你还能靠垃圾程度筛选；而它们全都长得一样好，于是你只能靠别的東西筛，而那个别的東西不在这一百个里面。

生长力有三条判准：生长性（往下还能长）、差异性（长出来跟别人不同）、纠缠密度（它牵动的东西多）。三条都不能在它长出来之前被验证。你只能在投入了之后知道。这就是为什么“选点子”不是一个筛选问题，是一个下注问题——而机器不下注。

上限三·不自动改姓。

它给你的东西带着来源的口音。一个从管理学语料里长出来的点子，说的是管理学的话；你要把它用在临床上，它一开口就露馅。

改姓不是换术语。它要深到概念语法那一层：这个领域的人怎么提问、什么算证据、什么算不言自明。一个没有改姓的点子，在目标战场上会被当场识别为外来户，而外来户的判决是不听。

而机器为什么不自动改姓？因为改姓的前提是知道目标战场的本地口音，而口音不在语料里——口音是那个领域的人不写下来的东西。

上限四·无长程一致性。

今天它给你一百个。明天你再问，它给你另一百个。这两百个之间没有关系。

而创新需要长程一致：一个东西要在几个月里被反复推、被现实打回来、被改、再推。这条线是一个人身上的东西——是他被这件事烦了三个月的那个连续性。机器每次对话都是新的，它没有被烦过。

上限五·不生成问题空间。

这是五道里最狠的一道。

它只在你给的问题里回答。你问“怎么提高这个渠道的转化率”，它给你二十个方法，每一个都合理。它不会告诉你：这个渠道的增长是补贴买的，还剩四个月，转化率是个假问题。

不是它不敢说，是它没有可以站的位置。它站在你给的问题里面，而“这个问题问错了”这句话必须从问题外面说。

五道上限里，一、四、五这三道在原理上不属于生成——它们要的不是更好的生成，是别的东西：一个问题空间、一段连续的被烦、一个能站在问题外面的位置。这就是为什么“等模型更强一点”救不了它。

三、六道工序，机器只压了第三道

把创新拆成可执行的工序，是六道：

抓核 → 抓裂缝 → 重组 → 改姓 → 锻造 → 投放

抓核：在一堆可以做的事里，认出那个值得押上去的核。三条判准（生长性、差异性、纠缠密度），都不能事前验证。

抓裂缝：找到旧语言接不住的那个地方——不是"还没人做过"的空白，是"大家一直站着的那块地板下面"。

重组：让散乱的材料"不得不一起被看"。这一道，机器很强。

改姓：换掉表层的壳，保住深层的机制，让它在目标战场上像个本地人。

锻造：六打——定义、论证链、反对意见、例子、方法、失效条件。把一个说法打成一件东西。

投放：选战场，扔出去，挨打，把反馈收回来接到下一轮的裂缝扫描上。

机器压缩的是第三道。它也能帮上第一、二道的检索部分（把地图铺开），能帮第五道的一部分（生成反对意见的初稿）。而抓核、改姓、投放，它一点忙都帮不上。

现在做那个算术。

六道工序的时间比例，大致是：

抓核 3：抓裂缝 4：重组 1：改姓 2：锻造 5：投放 1

>

总计 16 份。

机器把"重组"那 1 份压到接近 0。

16 → 15.01。提速约百分之六。

而所有人以为它提速了一百倍。

这个百倍是真的——在那 1 份里面，它确实快了一百倍。一个下午一百个点子，这是真的。问题是那 1 份只占十六分之一，而且是六道里最轻的一道。

这就是那公司的全部谜底：它把一道占 6% 的工序提速了一百倍，然后期待总产出翻倍。

四、可压的与不可压的

那另外十五份能不能也压掉？

能压的部分，机器已经压了，而且压得很干净：文献检索、数据整理、变体生成、格式排版、术语残留的机械检测。这些是纯赚，一分不用犹豫。

不能压的有三样，而它们恰好占了十六份里的十二份：

其一，沉浸时间。抓核和抓裂缝要的是一个人在一件事里泡够久。泡久了才有那个"哪里不对"的感觉，而那个感觉是裂缝的第一形态。你不能读一千篇论文来代替它——机器读一万篇也不会觉得哪里不对，它没有可以被硌到的地方。

其二，发酵时间。一个抓到的核，要在暗处待一段时间才能长成形。这段时间的外部表现是零产出。任何一个以周为单位交付的组织，在结构上取消了发酵。

其三，打磨时间。锻造那 5 份是全六道里最重的，而它没有任何捷径：定义要改二十遍，论证链要被人捶，例子要换三批，失效条件要自己写出来——而最后这一件是机器绝对做不了的，因为写失效条件的意思是"我承认我这个东西在这些情况下不成立"，那是一个承受动作。

三样不可压的时间，占十六份里的十二份。而它们不可压不是因为技术不够，是因为它们要的是一个人在场的时长，而在场不能被加速。

五、复合创新体：正确的分工

那分工该怎么划？只有一条原则：

人负责主导性问题空间与抓核定向，机器负责高频生成与变体试探。

具体到六道：

| 工序 | 人 | 机器 |

|---|---|---|

| 抓核 | 全部 | 把候选摆出来（不排序） |

| 抓裂缝 | 全部（那个"哪里不对"只能他有） | 铺开地图；列出久拖不决的论争 |

| 重组 | 指定方向 | 主力 |

| 改姓 | 全部（口音不在语料里） | 机械检测残留术语 |

| 锻造 | 主力 | 生成反对意见初稿、找反例 |

| 投放 | 全部（要挨骂的是他） | 整理战场信息 |

这张表里最要紧的一格是"抓核"那一行的右列：把候选摆出来，不排序。

一旦机器给了排序——"我建议第三个方案"——抓核当场蒸发。因为人接下来做的不是抓核，是审核一个推荐，而审核一个推荐的默认动作是通过。你把一百个摆平了不标高低，他才必须自己动脑子；你标了一个最优，他只会去检查这个最优的论证有没有毛病。他在验算，不在抓核。

第二要紧的是"锻造"那一行：机器生成反对意见，是这张表里性价比最高的一格。因为一个人对自己的东西天然看不见反对意见——而机器没有立场，它不心疼你的核。让它把最难听的话先说一遍，比等审稿人说便宜得多。

六、六道工序走一遍：一个可以点开看的例子

抽象的工序容易点头，难执行。这一节拿一篇真实的、读者可以点开看的东西走一遍——本站期刊论文频道的《[可规格化性：能力本位课程的隐蔽筛选器及其自我取消](#)》。选它是因为它六道全走完了，包括最容易被跳过的那道。

第一道·抓核。

候选很多：AI 与教育评价、AI 与教师角色、AI 与教育公平……每一个都是热点，每一个都能写。抓核抓的是"能力本位课程的目标表述"这一个点。

凭什么是它？三条判准：生长性——从它往下能长出课程、评价、政策三条线；差异性——所有人都在谈 AI 与教育，没有人在谈目标表述这个技术动作；纠缠密度——它牵动泰勒、马杰、布卢姆、核心素养、学业质量标准，一牵一大串。

这一步机器帮不上：它可以给你一百个候选，它不知道哪个的纠缠密度高，因为纠缠密度是相对于你要撬的东西而言的，而它不知道你要撬什么。

第二道·抓裂缝。

不是找空白（"AI 对课程目标影响的研究尚待深入"——这是空白，一钱不值）。是找那块所有人站着的地板："课程目标应当写清楚"——这句话六十年没人问过。而它下面埋着的是：写清楚 = 说得出口算做到 = 机器所需的全部条件。

这条裂缝的位置上堆满了文献：泰勒、马杰、布卢姆、整个能力本位改革。它不在空白处，它在最密的地方。

第三道·重组。

把三样从不放在一起的东西摆到一张桌上：测量学的信度要求、工程学的规格说明、生成模型的工作方式。三者一起看，那个等式就自己跳出来了——评分标准和规格说明是同一份文本的两次使用。

这一道机器很强。让它把这三个领域的定义并排列出来，二十分钟的事。而它不会主动把它们摆到一起——是人指定了要摆哪三样。

第四道·改姓。

那篇论文正文里，一个 SDE 术语都没有。它说的是泰勒的话、马杰的话、比斯塔的话。"可规格化性"这个词是在教育学的母语里当场造的，它接的是这个领域自己的文献。

为什么必须改？因为它要投的是教育学的期刊，而一篇需要先教会审稿人一套新话才能被读懂的论文，会在第一页就被判为外来户。

这一道机器完全帮不上：口音不在语料里。它能检测残留术语（机械活儿），它写不出本地人的那种句子。

第五道·锻造。六打，最重的一道，那篇论文的大部分时间在这里：

定义（可规格化性说到底是什么）；**论证链**（可测量→说得出判定条件→规格→可自动化的上界）；**反对意见**（表现性评价能不能救、隐性课程能不能扛、是不是只是技术局限——三条各自正面接住）；**例子**（核心素养框架是否逃脱）；**方法**（四象限、可识别性门槛）；**失效条件**（射程限于知识型目标、上界不等于充要、两难无解、最大风险是被滥用）。

最后这一打，机器绝对做不了。“我承认我这东西在这些情况下不成立”——那是一个承受动作，而它没有可以承受的位置。

第六道·投放。

战场：CSSCI 教育学。形制随之而定——中英双摘要、中图分类号、GB/T 7714 参考文献、与本领域实际论争对话。这些不是格式，这是战场的入场规矩。

而投放的真正含义是挨打：它会被审稿人捶，“可规格化性只是可编码性的换词”这一条会被捶得最狠——那一捶收回来，接到下一轮的裂缝扫描上，才算走完一圈。

现在回头看那个算术。这一圈里，机器的位置是：第三道全包，第一二道的检索部分，第五道的反对意见初稿。其余全部是人。

一个下午一百个点子，能给出的是这六道里的哪一道？第三道的一部分。而它连“该摆哪三样”都不知道。

七、这解释了一个悖论

本文的框架顺带解释了一件正在发生、而没有好解释的事：

为什么点子的产量暴涨，而“这个点子我早就想到了”这句话的出现频率也在暴涨？

因为一百个点子里，大概率包含了那个对的方向——机器确实把那个方向说出来了。于是每次有人做成一件事，总有人能翻出自己三个月前的对话记录，里面就有这个点子。

他没有说谎。那个点子确实在那儿。

而他没有做成任何事。因为：他没有抓核（他不知道这一百个里哪个是核），没有抓裂缝（他不知道这个点子撬的是什么），没有改姓（他的版本一开口就是外来户），没有锻造（他没有为它写过失效条件），没有投放（他没有为它挨过骂）。

“这个我早就想到了”——这句话在今天的准确含义是：那 1 份我有，那 15 份我没有。

>

而在点子昂贵的年代，这句话是有分量的，因为那时候有那 1 份就意味着你走过了那 15 份的一部分——它们本来长在一起。

机器把它们拆开了。而语言还没跟上：我们还在用“想到”这个词，指一件已经不再包含后面十五份的事。

八、这套东西什么时候失效

其一，六道工序这个划分本身可能是一张画错的地图。

凭什么是六道？为什么不是四道或者九道？本文没有论证这个划分的穷尽性与互斥性，它是从实践里归纳出来的，而归纳出来的东西通常带着归纳者的路数。如果这六道分错了，那么第三节那个 16 份的算术就是在一个错误的分母上做的——而那个算术恰恰是本文最有力的部分。这一条是本文最实质的弱点。

其二，时间比例 3:4:1:2:5:1 没有来源。

它是一个经验估计，不是一个测量。如果真实比例是 1:1:8:1:4:1——即重组其实占了一半——那么机器带来的提速就是 50% 而不是 6%，本文的整个结论要重写。这个比例是可以被测的（记录一个团队在每道工序上的实际投入），而它没有被测过，包括我自己。

其三，五道上限中的第二、三道可能会被攻破。

生长力判别和改姓，本文说机器做不了。理由是它们要的东西不在语料里。而如果模型开始接入一个组织的完整历史（所有的会议、所有的失败、所有的口音），这两道可能会松动。本文的判断建立在“口音不被写下来”这个经验事实上，而这个事实正在变化——当一个组织的每一次对话都被转写、每一次失败都被复盘并存档，口音就开始进入语料了。

其四，本文可能只是在为慢辩护。

“沉浸、发酵、打磨不可压”——这句话对任何一个做得慢的人都很受用。它可以成为一切拖延的护身符。本文给的分界（可压的已经压了，不可压的占十二份）是唯一的防线，而这道防线的两侧，在现实中经常分不清楚：一个人在发酵，和一个人在拖，从外面看是同一件事，从里面看也常常是。

回到那家公司。

它的工作坊现在一个下午产出一百个点子。它有一个专门的库存起这些点子，库里已经有六千个。每季度复盘时，会有人说“我们不缺点子”。

这句话是对的，而且它一直是对的——它在两年前就是对的，那时候库里只有三十个。

我们花了两年，把六道工序里最轻的那一道，从一个下午压到二十分钟。

然后我们站在六千个点子面前，等着创新自己发生。

>

而那六千个点子里，大概有二十个是真的核。它们和另外五千九百八十个长得一模一样，因为写它们的那个东西，从来没有为其中任何一个挨过骂。