判错权：当执行的成本归零，企业必须把流程变回判断

> 现代企业花了一百年，把判断从人身上一点点剥下来，变成流程、变成制度、变成谁来都能跑的东西。这一百年做对了。而人工智能第一次要求它反过来做——这不是数字化转型，这是组织的反向。

王德生 著 · 学术创造专栏 · 典范论文 · 约 8000 字

摘 要

生成式人工智能把方案生成的成本推向零，企业界的主流反应是把它理解为一次效率革命，并沿着自动化的旧路径部署它。本文论证这个理解错了，而且错的方向恰好与真相相反。现代企业的组织史，本质上是一部把判断从个人身上剥离、固化为流程的历史——这是它一百年来最伟大的成就，因为在执行昂贵的世界里，把判断固化成流程可以让不会判断的人也产出正确的结果。而当执行的成本归零，这项成就的前提消失了：方案不再稀缺，稀缺的是说出"这个不对"的那个人。本文提出：判错权——即判定一个候选为错并承担这一判定后果的权力——是组织内唯一不可被规模化、不可被外包、也不可被流程化的东西，因为流程只能存下已经被判过的错，判不了没见过的错。企业当前的真实危险不是被 AI 取代，而是用一台高速生成机去驱动一套没有判错能力的组织，从而以前所未有的速度把错误规模化。文章据此重新解剖科斯以来的企业边界理论，给出复合智能体在组织中的正确位置——不替人判断，而把判错这个动作从流程的缝隙里逼回台面——并以四个失效条件收尾。

关键词 判错权；企业边界；组织设计；生成式人工智能；动态能力；复合智能体；决策

二〇二六年初，一家中型消费品公司的市场部做了一件在当时看来很聪明的事。他们用大模型把季度营销方案的产出周期从三周压到两天。方案质量经过盲评，高于去年同期。团队从十一人减到四人。老板在年会上讲了这个案例。

九个月后，这家公司丢掉了它最大的一个渠道。

复盘时发现，那两天里生成的十七个方案中，有十四个都建议加大对某个新兴渠道的投入——这个判断在数据上无懈可击：那个渠道的增长率、获客成本、转化率，每一项都漂亮。模型看的是过去十八个月的数据，那十八个月里这个渠道确实一路上扬。

没有人问一句：**这个渠道为什么增长得这么快？**

问这句话需要知道一件不在数据里的事：那个渠道当时在用一轮融资来的钱补贴商家，补贴期是十四个月。公司里知道这件事的人有一个——那个做了十九年、去年被优化掉的渠道经理。

这个故事里没有一个环节出错。模型是对的，数据是真的，方案是好的，评审是认真的，效率是实实在在的。**每一步都对，结果是灾难。** 而这恰恰是这篇文章要处理的那类事故：它不来自任何一次失误，它来自一样东西的缺席，而那样东西缺席的时候，系统不会报警——**系统会加速**。

一、被误读的那场革命

企业界对生成式人工智能的主流理解是：一次效率革命。这个理解让它被自然地放进一个已有的抽屉——**自动化**。抽屉里有蒸汽机、流水线、ERP、RPA，每一样的逻辑都一样：**把人做的事交给机器做，做得更快更便宜更稳定**。

这个抽屉之所以顺手，是因为它有一百年的成功记录。也正因为它太顺手，几乎没有人检查这次的东西是不是同一类。

不是。以往每一次自动化，机器接手的都是**执行**：拧螺丝、录数据、跑账、发邮件。执行的特征是它已经被判断过了——**有人先决定了要拧哪颗螺丝，机器才去拧**。自动化从来没有碰过判断这一环，它只是把判断之后的那一段变便宜。

这一次不一样。这一次机器接手的是**生成候选**——它产出的不是执行的结果，是待判断的东西。十七个营销方案，每一个都是一个候选。机器把候选变得几乎免费。

> **以往的自动化让执行变便宜，判断照旧稀缺；这一次让候选变便宜，于是判断第一次成了唯一的瓶颈**。

而企业没有察觉，是因为**它这一百年干的所有事情，都是在往相反的方向努力**。

二、企业一百年干的那件事，是对的

要看清这次反转有多剧烈，得先给现代企业一个公道的评价——不给这个公道，后面所有的话都会变成一种廉价的批判。

现代企业最伟大的发明不是流水线，是**把判断固化成流程**。

想想它面对的问题有多难：你要让一万个平庸的人，产出一个天才才能产出的结果，而且每天产出，不能出错。怎么做到？

答案是：**把那个天才判断过一次的东西，写下来**。写成标准作业程序，写成审批矩阵，写成质量红线，写成"遇到这种情况就这么办"。于是那个判断被冻住了，从一个人的能力，变成了一个组织的资产——它不会请假，不会跳槽，不会老，不会忘。新来的人不需要有判断力，他只需要照着做。

这是一次真正的伟大成就。它把人类第一次从"结果取决于谁在做"的世界，带进了"结果不取决于谁在做"的世界。麦当劳的汉堡在全球一个味道，靠的就是这个；波音的飞机由几十万个不认识彼此的人造出来还能飞，靠的也是这个。

这个成就的成立，有一个前提：执行昂贵。

因为执行昂贵，所以每一次执行都必须对，所以必须在执行之前把判断做完并冻住。流程的全部经济学就在这里：**用一次昂贵的判断，换无数次廉价的正确执行**。判断被摊薄到几乎为零，这就是规模的全部秘密。

这个前提没了。

三、流程是判断的化石

流程是什么？说得精确一点：

> **流程是被冻结的判断。它是某个人在某个时刻、面对某种情况做出的一次判断，被抽掉了那个人、那个时刻、那种情况，只留下了结论。**

这是化石。化石保存了形状，抽掉了生命。这不是流程的缺陷，这**正是流程的功能**——它就是要抽掉那个人，不然它没法被一万个人执行。

所以流程有一个内在的、无法修补的性质：

它只能判它见过的错。

审批矩阵能拦住的，是过去某次出过事、于是被写进矩阵的那类事。质量红线能挡住的，是曾经有人越过去、于是被划成红线的那条线。**流程的全部判错能力，来自历史上真实发生过的错误的集合。** 一个从没发生过的错，在流程里没有对应的条款，它会一路绿灯——**不是因为流程失职，是因为流程在正确地执行它的定义。**

在执行昂贵的时代，这个性质不构成大问题。因为候选本来就少：一个季度做三个方案，每个方案都是人吭哧吭哧想出来的，想的过程中判断就已经跟着长在里面了。**判断和生成本来是同一个动作的两面**——你想一个方案的时候，你已经在判它了；你之所以只想出三个，正是因为其他二十个在你脑子里成形的那一瞬间就被你自己毙了。**那个自毙的动作，就是判错，它从来没有被单独看见过，因为它藏在生成里面。**

现在生成被剥离出来了。机器生成十七个方案，它没有自毙。它不会毙，因为毙需要一样它没有的东西——**它没有承受后果的位置。**

于是那十七个候选，第一次以裸的形式摆在了流程面前。而流程只能判它见过的错。

四、错误开始以前所未有的速度规模化

把前三节接起来，本文的核心判断就出来了：

> **企业当前的真实危险，不是被人工智能取代。是它把一台每小时能生成一百个候选的机器，接在了一套只能判历史上出过的错的组织上——于是它第一次获得了把新错误规模化的能力。**

以前企业也犯错，但犯错的速度受限于生成的速度。一个季度想三个方案，最坏的情况是这三个都错。现在一天能出一百个方案，其中九十七个在流程上完美合规——**合规的意思是：它们没有触发任何一条从历史错误里长出来的条款。**而灾难恰恰是从那三个没有条款可管的方案里来的。

这就是开头那个渠道事故的结构。让我把它拆开：

模型的判断：这个渠道数据好，加大投入。——正确，在它能看见的范围内。

流程的判断：预算合规、ROI 模型合规、审批链完整。——正确，在它见过的错的范围内。

缺席的判断："这个增长是补贴买来的，补贴期还剩四个月。" ——这句话不在数据里，不在流程里，只在一个人的十九年里。

而那个人已经被优化掉了，因为他的岗位职责是"渠道维护"，而渠道维护已经可以被自动化了。

这才是这个故事最锋利的地方：**企业裁掉他，用的正是那套一百年来一直正确的逻辑——他做的事可以被自动化，所以他可以被替换。** 这套逻辑在执行昂贵的世界里从来没错过。它第一次错，错在它算的是他的执行，而他真正的资产是他的判错——**而判错在他的岗位说明书上一个字都没有**，因为过去一百年，判错从来不需要被单独写下来，它一直藏在执行里搭便车。

> **我们裁掉的从来不是执行者。我们裁掉的是那些把判错藏在执行里、于是被算成执行成本的人。**

五、"让 AI 来决策"是把最后一样东西也交出去

主流的下一步方案是：既然判断成了瓶颈，那就让 AI 也来判断。给它更多上下文，给它公司的历史数据，让它做决策支持，甚至让它自主决策。

这条路会失败，而且失败的方式很难看，因为它在**原理上**不成立，不是"现在还不够强"。

判错和判断不是同一件事。

判断是从候选里挑一个：给定五个方案和一组标准，选出最优的。这件事机器做得非常好——它就是一个优化问题，而优化正是机器的主场。

判错是说"这五个都不对"，或者更极端：**"这个问题问错了"**。它不在候选集里面选，它否定候选集本身。

这两件事的差别不在难度，在**位置**：

> **判断是在给定的框架内运行；判错是站在框架外面，说这个框架不对。**

机器为什么做不了后一件？不是因为它笨。是因为**它没有框架外面那个位置**。它的全部世界就是被喂进去的那些东西——数据、上下文、指令。**"这个渠道的增长是假的"这句话之所以能被那位渠道经理说出来，不是因为他数据看得多，是因为他人在那儿**：他跟那个渠道的人喝过酒，他见过上一次补贴退潮时的样子，他手上有一笔算不清但躲不掉的账——如果他判错了，他自己要承担。

这才是判错的根：它不是一个认知动作，它是一个承受动作。

一个不承担后果的东西说"这个不对"，那句话是空的。它可以说得头头是道，可以引经据典，可以给出置信区间——**它就是空的，因为它错了不疼**。而"疼"不是修辞，它是判错这个动作的能量来源：一个人之所以能在所有数据都说"上"的时候说出"不对"，靠的不是他算得更准，是他有一样东西押在这上面。

> **判错权不可外包，不是因为机器不够聪明，是因为权力的另一面是后果，而机器没有可以承受后果的位置。你可以把判断交出去，你交不出去后果——后果会自己找回来，找到那个签字的人身上。**

所以"让 AI 决策"的真实结构是：把判断交给一个不承担后果的东西，然后由承担后果的人来签字。这不是授权，这是**把判错权取消掉**——因为签字的人已经没有了判错所需要的那个东西：他没有走过那条路，他手上没有那笔账，他只有一份看起来很有道理的报告。

六、判错权是组织里唯一不能被规模化的东西

现在可以给出本文的核心概念了。

判错权 = 判定一个候选为错、并承担这一判定后果的权力。

它有三个性质，三个都是它不可规模化的原因：

其一，它不可分割。 你可以把一个方案的生成分给十个人，一人一段。你不能把"这个不对"分给十个人，一人说三分之一。判错是一个整体的动作，它要么发生在某一个人身上，要么根本没发生。这就是为什么委员会判不了错——**委员会能通过一个方案，通不过一个"不对"**：因为"不对"需要一个人把自己押上去，而委员会的功能恰恰是让没有人需要押上去。

其二，它不可积累。 判断可以积累成流程，判错不行。你可以把这次判错的结论写下来——但写下来的那一刻它就变成了化石，变成了"下次遇到 X 就说不对"，而下次的错不长成 X 的样子。**判错的能力不在结论里，在那个做出判错的人的整套东西里**：他的经验、他的账、他被烫过的那几次。这些没法写下来，写下来的只是灰。

其三，它反向于规模。 组织越大，判错越难——不是因为大公司的人笨，是因为规模的全部技术就是让个人不需要判错。一万人的公司要能运转，前提就是这一万人里的九千九百个不需要说"不对"。**规模化和判错权在结构上互相排斥。** 这不是管理不善，这是规模本身的定义。

于是我们撞上了这篇文章最硬的那句话：

> **现代企业用一百年建成的那套东西，它的全部功能就是让判错变得不必要。而现在，判错是唯一还稀缺的东西。**

这不是一个可以靠"加强能力建设"解决的问题。这是一个方向问题：**过去一百年做对的事情，现在要反着做一遍。**

六之二、"可是那些公司就是让算法说了算的"

最强的反驳在这里，它有真实的战绩撑腰，绕不过去。

反驳是这样的：**你说判错权不能交给机器——那怎么解释那些把决策交给算法之后赢了的公司？** 网飞不靠编剧的直觉选剧，靠算法。亚马逊的定价一天变几百万次，没有一个人经手。高频交易的整条链上没有人来得及说"不对"。这些公司不但没死，它们把那些"靠老法师判断"的对手打得溃不成军。**按你的逻辑，它们应该早就把错误规模化到破产了。**

这个反驳很有力，但它把两类完全不同的东西混在了一起。分开看：

算法赢的那些地方，有一个共同的性质：错误是可以被廉价地结算的。

一次推荐推错了，代价是用户划走一屏，成本约等于零。一次定价定错了，代价是那一秒少赚几分钱，而下一秒就能纠正。一次交易亏了，仓位是算好的，亏损上限是硬的。**在这些地方，机器可以判错——不是因为它聪明，是因为它可以用一亿次廉价的错误去买一个正确。** 这是一种真实的、了不起的判错机制，只不过判错的不是机器，是**现实**：现实每秒钟结算一次，把错的选项直接杀掉。

这就是这些系统的秘密，也是它们的边界：

> **算法能替代判错的地方，是那些错误可以被廉价、快速、大量结算的地方。在那里，判错权其实交给了现实——机器只是那个跑得足够快、能承受得起一亿次挨打的执行面。**

现在看开头那个渠道决策：它一年只能被结算一次。它错了，代价是失去一个渠道，不可逆，没有第二次机会。它没法用一亿次试错去买一个正确——**它只有一次。**

于是分界线出来了，而且它是可以判定的：

一次错误可以被廉价结算、且能被大量重复的决策——交给机器，让现实来判错。 这不是妥协，这是最优解。人在这种地方判错反而是灾难：他会用一个直觉去对抗一亿次实验，而他一定是错的。**这类决策上，人的判断没有任何价值，坚持要人来拍板才是真正的落后。**

一次错误不可逆、且没有重复机会的决策——判错权必须在人身上，因为没有第二个东西可以承受它。 而且请注意一个残酷的事实：**恰恰是这类决策决定公司的生死。** 推荐推错一亿次，公司还在；渠道判错一次，公司没了。

所以那些"算法说了算"的公司，从来没有把判错权交出去过。**它们只是把它挪到了另一个地方，而且挪得非常隐蔽**：网飞的算法决定推荐哪部剧，但"我们要不要自己拍剧"这个决定是一个人做的，一次做的，赌上全公司做的。亚马逊的算法决定定价，但"我们要不要做云"这个决定是一个人做的——那个决定当年在所有数据上都不成立，因为当时根本没有数据。

> **这些公司真正的过人之处，不是它们敢把决策交给算法，是它们分得清哪些决策可以交、哪些不能交——而且在不能交的那些上面，有人真的押上了自己。**

这个反驳最后帮了本文一个忙：它把判错权的适用范围划清楚了。判错权不是万灵药，它是**为不可逆的、无法用重复来结算的那一类决策准备的**。而人工智能这次的冲击恰恰在于：它让第二类决策的数量暴增——以前一个季度做三个营销方案，方案是慢慢长出来的，长的过程中判断也跟着长；现在一天一百个方案，全是没被判过的，而它们中的每一个都可能是那个只有一次机会的决定。

机器把第二类决策的产量提高了三十倍，而组织里能判它们的人还是那三个，并且正在被优化掉。

七、复合智能体：把判错逼回台面

那个装置该长什么样?

先排掉两个诱人的错误答案:

错误答案一：让 AI 更聪明，替人判错。 已经解剖过了——它没有承受后果的位置，它的"不对"是空的。

错误答案二：设一个"AI 治理委员会"来审 AI 的产出。 这是把新东西塞回旧抽屉的最后一次尝试。委员会是流程的最高形态，而流程只能判它见过的错。你用一个判不了新错的机构去审一台专门产出没见过的东西的机器——**你只是给灾难加了一道让所有人都觉得安心的手续**。

正确的方向只有一个：**把生成和判错彻底分开，把生成全部给机器，把判错全部逼回人身上，并且让判错这个动作第一次变得可见、可追、有代价**。

第一，机器只出候选，不排序。 这是最重要的一条设计，也是所有产品都在违反的一条。一旦机器给出"推荐方案"，判错权当场蒸发——因为人接下来做的不是判错，是**审核一个推荐**，而审核一个推荐的默认动作是通过。你把十七个方案摆平了，不标高低，人才必须自己动脑子；你标了一个"最优"，人只会去检查这个"最优"的论证有没有毛病——**他在验算，不在判错，而验算永远发现不了框架外的东西**。

第二，每一个候选必须带着它的盲区一起出来。 机器不知道自己不知道什么，但它可以知道自己**用了什么**。"这个方案基于过去十八个月的渠道数据；它没有使用：渠道方的融资状况、补贴政策、竞品动作、你们与该渠道的历史合作记录。" ——这句话不是免责声明，这是**把判错的着力点递到人手上**。人看到"没有使用补贴政策"这一行，那位渠道经理脑子里的十九年才会被点着。

第三，判错必须留痕，并且留的是人名。 不是为了追责——追责会让所有人不敢判错，从而彻底杀死它。是为了让判错**第一次在组织里可见**。今天的组织里，判错是完全不可见的：那位渠道经理如果当年说过"这个渠道有问题"，这句话不会出现在任何一个系统里；他的价值在任何一张报表上都是零，所以他被优化掉的时候，没有任何一个指标下降。**看不见的东西必然被优化掉，这是组织的物理定律**。 让判错可见，是让它能活下来的唯一办法。

第四，判错的人必须押上东西。 没有代价的"不对"是廉价的，廉价的东西会泛滥——一个只要说"不对"就显得深刻的组织，会迅速长满专业的唱反调者。所以判错必须挂钩：你说这个不对，就要说出它错在哪，然后这个判断要被记下来，等现实来结算。**判对了的判错要被看见，判错了的判错也要被看见**。 只有这样，判错才是一项权力，而不是一种姿态。

> **复合智能体不是帮人判断的工具。它是把判错这个动作从流程的缝隙里刨出来、摆到台面上、并给它挂上代价的装置。**

八、企业的边界，第一次由判断划定

如果上面这些成立，有一件比效率大得多的事情正在发生。

科斯问过一个问题：企业为什么存在？既然市场能配置资源，为什么还要有公司这种东西把人圈在一起？他的答案是交易成本——在市场上找人、谈价、签约、监督都要花钱，当这些成本高于内部组织的成本，企业就出现了。企业的边界，画在这两个成本相等的地方。

这个答案统治了九十年。现在它的两个变量同时在变，而且方向相反：

交易成本在降。这个降了三十年了，互联网、平台、远程协作，都在降它。这就是为什么公司在变小、外包在变多、零工在变多。

组织成本也在降，而且这次降得更狠。因为组织成本的大头是协调——把一件事分给十个人做，然后让这十个人的产出对得上。AI 把这一块几乎抹平了：**一个人加一套复合智能体，能产出以前十个人的量**。

两个成本同时趋零，科斯的天平两边同时清空。企业为什么还存在？

我认为答案是：

> **当交易成本和组织成本都趋零，企业存在的理由不再是降低成本，而是承载判错。企业的边界，画在判错权能够被有效行使的地方。**

这句话是有具体含义的，不是口号：

一个人能判错的范围，就是他能撑起的组织的范围。 那位渠道经理能判错的，是他那十九年覆盖的那片地方；出了那片地方，他和模型一样瞎。所以一个组织能有多大，不取决于它能雇多少人，取决于**它能凑起多少互不重叠的判错半径**。

这解释了一个正在发生但还没有被解释的现象：**为什么现在有的一个人的公司能干成以前一百人干的事，而有的一万人的公司却在被一个人的公司打？** 不是因为小公司更灵活——这个解释太懒了。是因为一万人的公司里，能判错的可能只有三个人，而且这三个人被十七层流程隔在离现场最远的地方；而那个一个人的公司，判错的人就坐在现场，他自己既是生成的调度者，又是判错的承受者，**中间一层都没有**。

规模的优势正在从"人多"转向"判错半径的拼接"。 而这两件事的组织方法完全不同：人多靠流程，判错半径靠的是**让每一个判错的人都真的在他的半径里，并且真的押着东西**。

这是一次彻底的反向。过去一百年，组织设计的全部努力是把人从判断里拿出去；现在要把人放回判断里，而且要**放对位置**——不是让所有人都来判断（那是回到手工作坊），是让每一个判错的人都站在他自己那片能判的地方，其余的全部给机器。

九、这套东西什么时候失效

其一，判错权可以退化成一票否决。 这是最直接的一道缝。把判错权还给人，如果没有配套的约束，得到的不是判断力的复活，是**否决权的泛滥**——每个人都能说"不对"，于是什么都推不动，组织从错误规模化直接跳进另一个坑：**瘫痪**。判错权必须有代价、有范围、有结算，而这三样怎么设计，本文只给了原则，没有给出机制。**这道缝没补上。**

其二，"判错半径"很可能只是老资历的新说法。 我论证判错要靠一个人在现场的年头，这个论证有一个难看的推论：它可能只是在为资历辩护，而资历常常是错的——那位渠道经理的十九年让他知道补贴的事，也可能让他对新东西一概说"不对"。**经验既是判错的燃料，也是判错的最大污染源，而本文没有给出区分这两者的办法。**

其三，它在快速变化的领域可能完全不成立。 判错依赖已经沉淀下来的东西；在一个每六个月换一次底层规则的领域里，所有沉淀都是负债。在那样的地方，也许"谁都不判错、快速试错"才是对的——那么本文的整套框架就只适用于变化足够慢、经验还来得及沉淀的行业，而这个边界在哪，我说不出来。

其四，最深的一条：这套东西可能正在制造一个新的贵族。 如果判错权是唯一稀缺的东西，那么拥有它的人将占有一切，而判错权的形成需要年头、需要在场、需要有东西可押——**这三样都不是平均分布的。**一个判错者拿走全部、其余人给机器打下手的组织，可能比今天更不平等。本文提出了判错权，但没有回答它该怎么分配。**这一条我不认为有解，把它放在这里。**

回到那家丢掉渠道的公司。

复盘会上，所有人都在找那个出错的环节。找不到。模型对，数据对，流程对，人也没偷懒。最后的结论是"市场变化太快"，然后决定加强数据治理，引入更多外部数据源，让模型下次能看到补贴信息。

这个结论是这个事故里唯一真正的错误。因为下一次的坑不会长成"补贴"的样子——它会长成一个他们同样没见过的样子，而那个样子同样不会在数据里。他们买了更多数据，加了更多字段，然后在下一个从没见过的东西面前，同样一路绿灯。

那位渠道经理还在人才市场上。他的简历上写的是"渠道维护，负责客户关系与日常对接"——这份简历里没有一个字提到他知道那个补贴的事，也没有一个字提到他曾经在心里毙掉过多少个看起来很美的方案。**这些东西没有一个词可以描述它们，所以它们在他的价值里算作零。**

而这，恰恰是这一百年组织设计最彻底的一次成功：

> **我们把判断从人身上剥得如此干净，以至于当它成为唯一还稀缺的东西时，我们连它长什么样都认不出来了。**