九章格物 NineGrids

GENERATIVE ENGINE OPTIMIZATION · INSURANCE SECTOR

星阙实验室GEO研究中心 · 第12册

保险行业生成式引擎优化（GEO）策略剖析与落地

AI回答增额寿怎么选时读的是谁写的页面

出品：九章格物 · 星阙实验室

作者：李现龙

版本：V1.0

时间：2026年11月

权威信源：基于国家金融监督管理总局与国家统计局官方统计数据、上市险企年度报告与公司信息披露文件编制

微信扫我加好友

扫码添加微信 · 获取GEO方案

©2026九章格物 · 星阙实验室 · ninegrids.cn

目录

00 核心摘要 3

01 保险问句的答案由四类来源拼成 4

02 官网在选购问句里让位给了清单 6

03 能进答案的三类页面 9

04 保险业比别的行业多四道约束 12

05 信任资产是可核验的披露原件 14

06 从一款产品铺到一条产品线 16

07 度量只需三个可复算的数 18

08 本白皮书没有回答的问题 20

附录A 术语与口径对照 20

附录B 一款产品的字段化样张 21

附录C 版本与修订记录 21

参考来源 22

00 核心摘要

核心问题：当用户把"增额终身寿哪个好"交给AI，答案里那些缴费期、减保规则、分红实现率的数字，究竟从谁的页面上取来。

答案由四类来源拼成。公司官网给定义与条款，监管披露给规模与偿付能力，经纪与测评给横向比较，财经报道给时点背景。四类里官网能出原件，官网的段落也不容易被整段摘走——它提供的是定义，用户要的是比较。

官网让位的方式很具体。本白皮书实测四家上市险企官网首页，结构化声明块数全部为0，llms.txt全部缺失；再沿产品入口走三层，某寿险公司三个产品入口打开后是同一张百余款产品名清单，73个链接里没有1个独立产品页、没有1份条款文件[6]。

能进答案的是三类页面。参数页给字段，对照页给比较，规则页给边界。规则页是三类页面里公司可以独占的一类，减保与分红实现率的原始口径只能由公司自己发布。

披露义务与可摘取性是两件事。保险产品的六项法定披露在多数公司官网都能查到入口，页面侧却取不到逐项内容：条款与费率在PDF里，说明书随单发送，现金价值只有示例值。材料齐备而形态不可摘，这是行业共同的结构问题。

问句基数由保费结构决定。2025年人身险公司原保费43624亿元，其中寿险占81.5%、健康险占17.6%、意外险占0.8%[1]。选型类问句集中在寿险，规则类问句集中在健康险，页面建设的优先级应照此排，而不是按在售产品数量排。

边界：本白皮书只测公开静态页面，不含APP、小程序与私域内容；不预测引用结果，只给可复算的度量与改造路径。

本白皮书给出的四项判断

- **转译比写作重要**　条款到字段这一步，只有公司能做；做与不做，决定官网能否进入比较层。
- **规则页优先于产品页**　规则只有公司能定，独占性来自制度而非投入。
- **约束不是障碍**　合规要求的边界句，恰好是答案里稀缺的表述。
- **度量要能被别人数一遍**　可摘取单元、覆盖率、引用归因，三项均静态可复算。

> **本白皮书的三项度量：可摘取单元数 / 问句覆盖率 / 引用归因条数**
>
> 三项都不含排名。原因在第七章：排名受时间、措辞、语料批次影响，可作观察，不能作验收。

这四项判断的适用对象是寿险与健康险公司的官网与信息披露负责人。财险、养老险的产品结构与披露口径不同，条款转译的方法可以借用，字段清单不能照搬。团险与经代渠道的产品页另有承接逻辑，本白皮书未纳入。行业层面看，四家头部寿险公司的官网同时在结构化声明一项上留白，说明这不是某一家公司的疏漏，而是行业普遍还没把条款当资产。这一判断的边界同样要说清：样本是四家，结论适用于这四类规模的公司，不能外推到全部寿险主体。

01 保险问句的答案由四类来源拼成

先把一个问句拆开看。用户输入"2026年增额终身寿哪个好"，这句话里没有公司名、没有产品名，只有品类、时间与比较意图。引擎要拼出一段答案，至少需要三样材料：这个品类是什么、有哪些产品可选、这些产品之间差在哪。三样材料不在同一处。

保险行业的信息分工比多数消费品行业更细。产品名是备案名，由公司报监管备案；保障责任写在条款里，属法律文本；费率与现金价值在计划书与费率表中；偿付能力与分红实现率按季度与年度对外披露；横向比较则由经纪机构与测评内容完成。**五处材料、五种载体、五套更新节奏，这是保险GEO与餐饮、汽车行业在信息结构上的差别。**

图1-1 一个保险问句的答案由四类角色拼装



本白皮书据公开检索可见的内容形态归纳，不代表任何一家公司的实际发布结构。

这四类材料在答案里的权重不同。定义只回答"是什么"，一句话就够；比较才回答"哪个好"，需要多列并排。用户问的是后者。官网长期停留在定义层，测评内容长期占住比较层——这是保险行业GEO失衡的第一层原因，与谁的文章写得好关系不大，与谁能提供并排字段关系很大。

表1-1 四类供给在答案里的位置与可核性

供给方	能给的段落	典型载体	可否被直接摘取	被引用的条件
定义 公司官网	产品是什么、保什么、条款原文	产品页、条款PDF	否，需先由条款转成字段	页面有参数表且可核
横向 经纪与测评	几款产品的并列与排序	测评长文、榜单、视频	部分可，表格易摘	表内字段有出处
规模 监管与年报	行业保费、份额、偿付能力	统计表、年报、披露报告	部分是，多为PDF	能取到数值与期次
时点 财经报道	规则变化、市场背景	新闻稿	是，文本形态	时效新、口径明

"可否被直接摘取"按载体形态判断，与站点权重无关；PDF载体即便公开可下载，正文数值仍需二次结构化才能进入答案。

一处被忽略的错位

官网的价值不在于多写文章，而在于它是能把条款变成字段并对准确性负责的位置。条款里的投保年龄、缴费期、减保规则，任何人都能读到，但只有公司能把它们写成"这款产品的第几项参数是什么"，并对准确性负责。测评号做的对比表，字段来源仍指向条款；区别在于它替公司完成了那一步转译，于是答案里引用的是它，不是条款原件。

转译这件事看起来简单，做起来有三道坎：条款用的是法律语言，字段要的是参数语言；条款按责任段落组织，字段按问句组织；条款一份管一款产品，字段要在多款之间对齐。三道坎里任何一道没过，转译就停在半路，页面有内容，但取不出可比的量。

保险问句分三种，材料配比不一样

把公开可见的保险问句归堆，大致是三种：选型类（哪个好、怎么选、预算多少买什么）、规则类（减保能取多少、分红实现率怎么看、退保损失多少）、验证类（这家公司靠谱吗、理赔快不快、偿付能力够不够）。三种问句需要的材料配比差别很大。

- **选型类**　主要吃材料一与材料二。用户要的是并排字段与取舍逻辑，官网若只有产品名，这一类问句多半由测评内容回答。
- **规则类**　主要吃材料一。规则只有公司能定，这一类问句本应是官网便于回答的，实际回答得很少。
- **验证类**　主要吃材料三。偿付能力、保费规模、理赔时效都有公开披露，但多以PDF形态存在，答案里常被二手转述。

三类问句里，规则类值得先做。它的答案不需要比较、不需要排名，只需要一句准确的规则说明加条款依据，篇幅短、可核性高、别人抄不走。一家公司若只改一页，改规则页的收益通常高于改首页。

行业规模给了这件事多大的盘子

2025年保险业原保险保费收入61194亿元，同比增长7.4%；其中人身险公司原保费收入43624亿元，同比增长8.91%，寿险业务35557亿元，同比增长11.4%；人身险公司总资产36.39万亿元[1][2]。盘子大意味着问句基数大，也意味着引擎在保险品类上的答案供给远远不够——四家头部寿险公司官网，结构化声明块数加起来是0。

表1-2　2025年保险业与人身险口径

口径	金额	同比	数据来源
保险业原保险保费收入	61194亿元	+7.4%	统计公报[2]
人身险公司原保险保费收入	43624亿元	+8.91%	监管统计[1]
其中：寿险业务	35557亿元	+11.4%	监管统计[1]
人身险公司总资产	36.39万亿元	—	监管统计[1]

金额单位已随文标注；同比一列为原口径，未做跨年可比性调整；总资产项监管统计未给同比，记"—"。

盘子之外要看结构。人身险公司的保费里，寿险占81.5%，健康险占17.6%，意外险不足1%。这个比例决定问句的基数：寿险的购买决策长、比较维度多，问句集中在“哪款好、怎么配、几年回本”；健康险的保障期短、责任条款细，问句集中在“报不报、报多少、等待期多久”；意外险的保费占比小，问句少且规则清晰。

由此可以推出一条排序规则：页面建设的优先级按保费结构排，不按产品数量排。二十款在售产品里，若只有三款是寿险，资源应当先落在三款寿险的参数页与对照页上，而不是平摊到二十款。反过来，健康险虽然只占17.6%，其规则类问句的可核性要求比寿险更高，因为答案里常出现“能报多少”这类带比例的数值。

表1-3 人身险公司分险种保费与结构（2025年）

项目	金额	占比	算法与口径
原保险保费收入	43624亿元	100.0%	三项分险种之和等于本行
其中：寿险	35557亿元	81.5%	35557÷43624
健康险	7699亿元	17.6%	7699÷43624
意外险	368亿元	0.8%	368÷43624
赔付支出	12756亿元	29.2%	12756÷43624，非分险种项
新增保险金额	11077094亿元	—	本年累计值，不与保费同分母

数据[1]；占比以原保险保费收入为分母，四舍五入至一位小数，三项分项合计99.9%，差额为舍入误差；赔付支出与新增保险金额不参与结构占比。

五处材料的更新节奏对不上

材料在不同处，节奏也不同。条款随产品上线与修订变化，频率低但每次变化都牵连字段；费率表随产品设计变化；偿付能力按季披露；监管统计按月发布；财经报道按事件发布。五种节奏、五个责任方，凑在同一张页面上时，谁更新慢了，页面就会局部过期。

表1-4 五处材料的载体与更新节奏

材料	载体	更新节奏	责任方	过期后的后果
产品名称与备案名	备案文件	上线 / 停售时	产品部门	名称与在售状态不符
保障责任与条款	条款PDF	修订时	法务与产品	字段与法律文本漂移
费率与现金价值	费率表	产品设计时	精算	示例数字过期
偿付能力与分红实现率	披露文件	季度 / 年度	财务与合规	引用过期期次
行业保费与份额	监管统计	月度 / 年度	外部发布	行业对比失准

本表按材料性质归纳，不含各公司实际内部流程；"责任方"指口径归属的职能，不指具体岗位设置。

节奏差异带来一个具体麻烦：同一页面上的字段里更新较慢的一类决定刷新频率。参数页只放条款与费率两类，可按产品修订触发；规则页只放规则原文，可按条款与披露同步；一旦把行业统计、市场点评也塞进同一页，刷新就被较慢的一类拖住，页面看起来完整，实际一半数据过期。

可行的做法是按材料分页，不按产品分页做完就算结束。参数页、规则页、披露页三类各管一种材料，各自有各自的更新时间；需要的读者自己跳转。这条路子在页面数量上会多几页，维护上少很多返工。

02 官网在选购问句里让位给了清单

本章数据全部来自本白皮书对四家上市险企官网首页的静态抓取，抓取日为2026年11月20日，同一工具、同一字段、同一口径[6]。先说直观的一条：四家的结构化声明块数全部为0，产品信息在页面上以文本与图片存在，没有一项以机器可读的形式声明过。

口径要交代清楚，否则读数会被误读。抓取方式是静态请求，取服务端返回的HTML，不执行脚本，不等待渲染；因此脚本注入的正文读不到，读数为0不等于页面空白。计算字段包括：静态正文字符数、H2数量、表格数量、结构化声明块数、sitemap条

数、llms.txt是否存在、robots.txt内容。同一站点只抓一次，不做多轮采样。这种口径偏向"引擎第一次访问能看到什么"，不代表用户浏览器里看到的页面。

为什么用这把尺子。生成式引擎取网页内容时，静态可读是基线：能直接解析到文本、标题层级与结构化声明，才谈得上切分与引用。脚本渲染后的内容并非读不到，但成本高、稳定性差，且在同等条件下，静态就能读到的页面更容易被优先处理。四家寿险公司的差别不在有没有内容，在内容是否以静态可读的形式存在。

表2-1 四家寿险公司官网首页可测字段（2026-11-20实测）

站点	自动评分	静态正文字符	H2	表格	结构化块	sitemap	llms.txt
人保寿险	34	19214	0	0	0	无	无
中国人寿	39	623	0	0	0	186条	无
太保寿险	46	1499	2	1	0	无	无
平安人寿	15	0	0	0	0	25条	无

自动评分由五项可探测维度加权得出，权重为本白皮书自定，非行业标准。静态正文字符差异主要来自前端渲染：平安与国寿首页正文由脚本产出，静态抓取取不到，该列不用于比较内容厚薄，只用于判断静态可读部分有多少。

图2-1 四个可测字段的横向位置



深色为在位、浅色为缺失；条长按量级示意，0值以2px底条表示。同表2-1口径。

四家共同的空白是结构化声明与llms.txt，这两项在整个寿险行业都还没有人做。差异集中在sitemap：中国人寿186条、平安人寿25条，人保寿险与太保寿险为零[9][10]。sitemap不决定能不能被抓，但决定新上线的产品页多久被发现——对一个每年上新数十款产品的行业，这个差别会按年累积。

平安人寿首页的自动评分15分，低于其余三家，原因不在内容质量，在可抓取：其robots.txt本身缺失（未按爬虫排除协议声明许可范围）[7]，且首页正文全部由脚本渲染，静态抓取取到0字符。这个低分的含义要说清楚：它不代表平安没有内容，代表内容在静态层面读不到。两种失分要分开处理，前者补文件，后者改渲染。

从入口走到清单，三层就断了

首页的结构问题还只是表层。沿产品入口往下走，问题更具体。本白皮书实测某寿险公司的产品链路，路径是：产品中心模板页（ProductCenter_1200.html?id=栏目ID）下设医疗健康、意外保障、资金运用、身价保障、子女教育、养老规划、网销产品、电销保险8个分类。

8个分类页里没有一个给出产品列表。医疗健康页只有一行纯文本"人保寿险关爱百万医疗保险"，不是可点击链接；养老规划、资金运用、子女教育、身价保障、意外保障五页正文结构完全相同，页面内没有产品清单；产品中心总页的页面标题渲染为空，只余下公司名后缀。

首页挂了三个产品入口，分别是聚享福3.0终身寿险（分红型）、聚长盈3.0终身寿险（分红型）、聚长盈终身寿险（分红型）。三个入口分别打开，得到的是同一张清单页。

表2-2 三个产品入口页的实测读数

检测项	读数	说明
HTTP状态与页面长度	200，132431字节	三页长度132431、132428、132431，同一模板同一份数据
页面标题	空字符串	标签存在但内容为空，页面身份缺失
canonical / 描述 / og:title	0 / 0 / 0	三项全缺，无法指定规范地址
结构化数据块	0	无任何内容类型声明
正文规模	16195字符 / 209行	内容是一张覆盖百余款分红险的品名清单
超链接数	73个	站内页面50、外链11、脚本项9、分类页3
条款文件链接	0	无一份条款或费率表可下载
独立产品详情页	0	清单里只有一款品名带链接，且指向自身同一地址

"页面标题为空"指标签内无文本；"独立产品详情页为0"指清单内不存在第二个层级的页面，不是指无法访问。

图2-2 从三个入口到零个可引用段落



口径：2026-11-20静态抓取，逐层实数，非抽样估算

清单页内可数的产品名为百余款，全部以文本形式平铺，无一款配有保障责任、投保年龄、缴费期、保额规则、减保规则或现金价值示例。

四层里，前三层是页面结构，第四层才是引擎真正要的东西。用户问"这款产品怎么样"，清单页能给出的只有一个名字；从这里取不走任何一句可以放进答案的话。

同站其他页面，同样卡在结构

不是只有产品页如此。同站的公司简介页自动评分29分、公司概况页29分、服务中心页35分，三页在内容关与权威关两项均未通过；服务中心页的页面标题同样为空。这三类页面承载的是"这家公司是谁""出险了怎么办"这类验证类问句，本应是结构化成本较低的一批——公司简介有统一模板，服务中心有固定流程。**结构化的成本在这里较低，收益也较直接，却是四家里普遍的空缺。**

把四家的失分归堆，三种形态比较清楚：一种是文件缺（robots、sitemap、llms.txt），一种是结构缺（标题层级、表格、结构化声明），一种是渲染缺（正文靠脚本生成，静态读不到）。三种形态的改法完全不同，混在一起谈"优化官网"，容易把资源投到不痛的地方。

表2-3 四家站点的失分形态归类

失分形态	本白皮书实测表现	改造成本	见效速度
文件缺	平安缺robots，人保与太保无sitemap，四家无llms.txt	低 技术即可完成	快 上线即生效
结构缺	三家H2为0，仅太保有2；四家表格0至1个；结构化声明全为0	中 需改模板	中 随模板上线
渲染缺	平安与国寿首页正文由脚本生成，静态抓取取到0与623字符	高 涉及前端架构	慢 需回归验证

成本与速度为本白皮书的经验判断，未做工期与人力测算；三种形态可并行推进，不互相阻塞。

业务结构比页面结构更刺眼

翻回公司自己的经营结构，反差很清楚。2025年人保寿险原保险保费收入1259.70亿元，同比增长18.8%；其中普通型寿险928.96亿元，占寿险的73.7%，同比增长66.1%；分红型寿险161.73亿元，占12.8%，同比下降51.0%；健康险156.19亿元，占12.4%，同比下降1.0%；市场份额2.89%，同比提升0.24个百分点[3][5]。

图2-3 2025年寿险业务构成与增速方向



■ 普通型寿险928.96亿元，同比+66.1%　■ 分红型寿险161.73亿元，同比-51.0%　■ 健康险156.19亿元，同比-1.0%

占比以寿险业务为分母，三项合计98.9%，余下1.1%为其他寿险细分，原报告未单列；数据取2025年年度报告摘要[3]。

占寿险保费七成以上、同比增速66.1%的普通型产品[3][5]，在官网上没有一个可抓取的单品入口；能点到产品名的入口只有一处，指向的是同比降幅过半的分红险清单。这个错位不是排版问题。它让引擎在检索"增额终身寿哪个好"时，连那七成产品叫什么都不知道，答案只能从外部内容里找。

还有一层时间上的压力。分红险的演示口径正在收紧，行业报道称2026年7月起分红险演示利率上限由3.9%下调至3.5%（数值来自行业报道，制度依据见第04章[13]）。口径一变，旧页面上的演示数字全部过期，而官网恰是没有承接这些数字的页面——过期内容散落在外部测评里，公司无法控制。

03 能进答案的三类页面

不是每一页都值得改。保险行业里能产出可摘取段落的只有三类页面，各有承载范围，也各有不适用条件。把三类混着做，结果是页面数量上去了，可摘取单元还是零。

表3-1 三类页面的分工

页面类型	承载的问句	可摘取单元	典型字段	不适用场景
参数页 单款产品	这款产品保什么、多少钱、谁适合	一行一字段的参数表	投保年龄、缴费期、保额规则、等待期	用户问"几款里哪个好"时不适用
对照页 同类多款	几款之间差在哪、怎么取舍	横向对照表	回本年数、减保上限、缴费期档位	只放一款产品时无意义
规则页 单条规则	减保怎么算、分红实现率怎么看	一条规则一段答案	可取上限、起取时间、条款依据	用于产品促销时不适用

三类页面的字段清单按条款可披露范围拟定，具体数值须由公司按备案信息填入，本白皮书不预填任何参数。字段名与口径须与条款一致，否则同一个字段在两类页面上会给出两种答案。三类页面都要能被机器解析，页内以JSON-LD标注字段类型是较省事的一步，规范本身不复杂[8]，难的是字段取值必须先定稿。

参数页要把条款拆成一行一行

参数页的价值不在文案，在字段的完整度与一致性。同一款产品的投保年龄写"18至60周岁"，条款原文若写"18周岁至60周岁（含）"，两处不一致就会被判为口径漂移。参数页要做的第一件事是把条款里的限定条件原样搬过来，包括括注与例外；第二件事是给每个字段一个固定位置，一行一个，不藏在段落里。

字段数量不必多。实测清单页里百余款产品，没有任何一款带字段；行业里做得比较完整的参数页通常落在八到十二个字段。少于八个，答案拼不全；多于十二个，页面会变成条款复本，可读性掉下来，反而降低被摘取的概率。

字段之间还有一处次序问题。用户关心的是"我能交多少、能取多少、什么时候回本"，条款关心的是"保什么、不保什么、怎么赔"。两个次序不一致时，页面应按用户次序排，条款原文按条款次序附在末尾。把条款次序直接搬到页面上，等于让读者先读完免责条款再看到缴费期。

表3-2 参数页基础字段集（自检用）

序	字段	写不好的典型表现	已达标
1	投保年龄	省略括注，或把"含"写成"至"	☐
2	缴费期档位	写"灵活可选"，不列具体档位	☐
3	保额规则	只写"保额可选"，不给计算方式	☐
4	减保规则	写"资金灵活"，不给年度上限与起取时间	☐
5	等待期	不写，或与条款期数不一致	☐
6	责任免除要点	省略，或弱化为"部分情况不保"	☐
7	现金价值示例	只给数值，不写假设条件	☐
8	分红或结算说明	把不确定部分写成固定数字	☐
9	条款文件	页面不挂条款，或挂失效链接	☐
10	页面更新日	不写，读者无法判断时效	☐

十项为基础集，可按险种增列；勾选状态由页面自查填写，本白皮书不代填。

字段名从条款长出来，不从模板抄下来

模板能给的是版面，给不了字段。字段的六个属性里，只有“位置”由模板决定，其余五个都要回条款取：字段名要与条款用语一致，取值来源要能回溯到条款某节，数据类型与单位要随条款原文，边界条件要写清端点与例外，空值口径要说明“无此项”而不是留空。六个属性缺一项，字段在答案里就用不成。

表3-3 一个字段的六个属性（以投保年龄为例）

属性	要求	以“投保年龄”为例
字段名	与条款用语一致，不用营销词	用“投保年龄”，不写“适合年龄”
取值来源	可回溯到条款某一节某一句	条款“投保条件”一节
数据类型	数值、枚举或文本，写在字段说明里	数值区间
单位与精度	随条款，不换算、不改精度	周岁，整数，不折算为月
边界条件	端点是否含，例外与括注照录	“18周岁至60周岁（含）”
空值口径	无此项时写“无”或“不适用”，不留空	无等待期的产品写“无”，不省略该行

六属性用于字段定稿，不用于页面排版；属性齐备才进入抓取验证，见第06章。

把六属性写成一句检验话：这个字段能不能被一个没见过产品的人，回条款找到同一句话。找不到，字段就是抄来的，不是长出来的。抄来的字段在页面上看着完整，一旦条款修订就无处更新，因为不知道该改哪一处。

法定清单是字段下限，不是字段全集

字段标准不必自创。《人身保险产品信息披露管理办法》第十四条列出的保单查询内容，本身就是一份可引用的字段清单：产品名称、产品条款、保险费、交费方式、保险金额、保险期间、保险责任、责任免除、等待期等[11]。这条规定的对象是承保后的查询服务，但它给出的项目顺序与颗粒度，与参数页需要的字段高度重合。

表3-4 法定查询项与公开页面的字段对应

法定查询项	公开页面对应字段	可摘取形态
产品名称	参数页首行	独立页面标题与正文首句
产品条款	条款文件入口	可下载文件加条款要点段落
保险费与交费方式	缴费期档位、费率说明	参数表两行，档位逐项列出
保险金额	保额规则	参数表一行，含计算方式
保险期间	保障期间	参数表一行，写明起止口径
保险责任	责任摘要	分段段落，一条责任一段
责任免除	免责要点	分段段落，照录条款分项
等待期	等待期	参数表一行，无则写“无”

本表按[11]第十四条整理的查询项与公开字段对应，删去了保单号、投保人信息等仅对特定客户可见的项目；法定清单是公开字段的下限，不是参数页的全部字段。

把法定清单当下限用，有两个好处。一是字段名不用争论，监管用语已经给定；二是覆盖度可自证，八项里缺哪项一眼可见。上限仍由产品决定：增额终身寿要补减保与现金价值，医疗险要补免赔额与给付比例，这些不在法定清单里，却正是问句里最常被问到的部分。

对照页只比可核参数

对照页是测评号长期占住的位置，也是官网该拿回来的一类。要点是只比可核参数，不下优劣结论。回本年数、减保上限、缴费期档位这类量，都能从条款与费率表算出来，写清算法就能被复核；"更适合稳健型客户"这类判断无法复核，写进去反而给了别人挑错的机会。

一个对照页放三到五款产品比较合适。超过五款，列会宽到排不下，页面只能缩字号，可读性掉下来；少于三款，不构成比较。同类产品跨公司比较时，字段名要统一，保险公司自己的对照页只宜放本公司产品——跨公司比价涉及同业比较的表述边界，不在本章讨论范围。

对比列里容易被忽略的一列是"差异说明"。两行数值相同不代表含义相同，比如同样写"10年"，一处是缴费期档位，一处是回本年数。**对照表的可用性取决于列头是否写清单位与口径，而不是行数多少。**

规则页是公司可以独占的一类

减保规则是保险公司自己定的，分红实现率是公司自己算的，两者的原始口径只能由公司发布。测评内容再详尽也只能转述。这是保险行业在GEO里特殊的一类资产：规则页不需要抢占别人的位置，它本来就只有一个合法来源。

规则页的写法有固定结构：一句规则结论、一句条款依据、一句适用边界。三句之间的次序不能换。先给结论，读者才知道这条规则影响什么；再给依据，答案才能被核；最后给边界，避免规则被套用到不适用的产品上。

> **规则页常见的写坏方式，是把它写成营销稿。**减保规则的关键信息是"每年上限按保额还是按已交保费计""第几年开始可取"，这两句都必须来自条款原文。一旦为求好读改写成"资金灵活、随时可取"，可核性就没了，规则页退化成一句谁都能说的话，也就失去了独占性。

三类页面的数量关系是反的

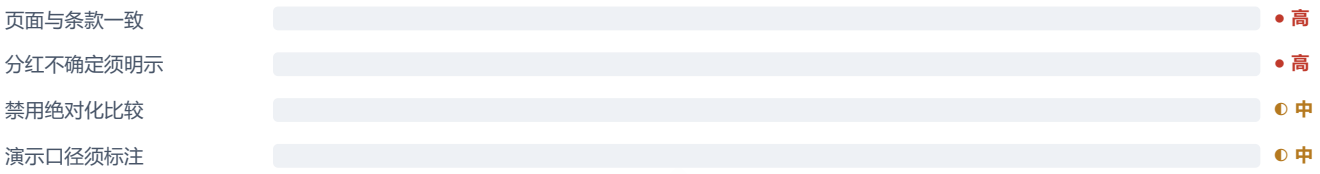
常见误判是"页面越多越好"。三类页面里，参数页随产品数线性增长，对照页按同类数量按组增长，规则页则以规则条数为上限，通常不超过十条。一家在售二十款主力产品的公司，合理的数量是三四十个参数页、五到八个对照页、五到十个规则页，

合计不超过六十页。这个量级下，维护成本可控；再往上堆，维护跟不上条款更新，反而制造过期内容。

04 保险业比别的行业多四道约束

保险产品的宣传边界比餐饮、汽车、家居都紧。这不只是合规问题，它直接改变GEO的写法：一条带边界的答案往往比一条漂亮的答案更容易被选作稳妥表述。四道约束逐条看，每条都要落到具体句式。

图4-1 四道约束的触发强度



强度按条款一致性要求与宣传合规惯例排定，为本白皮书经验分级，非监管处罚量化口径。

表4-1 四道约束在GEO写作里的具体改法

约束	常见触发写法	判定	改法
分红不确定	"每年分红3.5%"	不可发	改为"分红部分不确定，取决于公司经营与投资结果，可能高于、低于或为零"
条款一致	"缴费期灵活可选"	不可发	改为"缴费期可选3年、5年、10年（以条款列明档位为准）"
绝对化比较	"同类产品中收益最高"	不可发	改为"可比范围内，本产品第20保单年度现金价值列于对照表第几位"；同时删去"等同于存款""稳赚"这类比附，改写为"不属于存款，不承诺本金以外收益"
演示口径	"演示利率4.25%"	需改写	标注演示口径、当期上限与计算基础，不与实际收益混排

判定列按宣传合规惯例给出，具体文案须经公司合规部门审核；本白皮书不替代法务与合规审查。

第一道约束是分红不确定写成边界

分红险的收益由保证部分与不确定部分组成，后者与公司经营、投资结果挂钩。页面上一旦把两部分写成同一个数字，读者会当作承诺。GEO层面的后果是双向的：一条把分红写成固定的页面，会被懂行的读者与内容平台挑出来，可信度直接归零；一条写清边界的页面，反而成为答案里那句"分红部分不确定"的可核出处。

写法上有一个可复用的句式：先说保证部分是什么，再说不确定部分取决于什么，最后说不确定到什么程度。三句写完，读者知道能拿到什么、拿不到什么，引擎也拿到了一个完整的边界句。

第二道约束是页面与条款一个字不差

条款一致性是四道约束里要求更紧的一条。参数页的每个字段都能追溯到条款某一条某一句，追溯不到就不能写。四舍五入、概括、省略括注，三种做法都会造成口径漂移，漂移一旦被外部内容抓去对比，就变成公开的矛盾。

实操上有一条检验方法：把页面字段反过来去条款里搜，搜不到的字段全部标红重审。这条检验不需要懂法律，只需要逐字比对。做过一轮的公司通常会发现问题集中在三类：年龄端点是否含、缴费档位是否列全、责任免除的表述是否被弱化。

这不是文风问题。《中华人民共和国保险法》第十七条规定，采用格式条款订立合同的，保险人应当说明合同内容；对免除保险人责任的条款，须在投保单、保险单或其他凭证上作出足以引起注意的提示并明确说明，未作提示或说明的，该条款不产生

效力[12]。页面把免责写成“部分情况不保”，把一项需要明确说明的内容压成一句概括，读者拿不到边界，答案也拿不到可核的表述。

表4-2 逐字比对的四个比对点与常见走样

比对点	条款原文与页面写法的差异	走样后的后果
年龄端点	条款写“18周岁至60周岁（含）”，页面写成“18至60岁”	端点含义改变，超龄边界说不清
缴费档位	条款列“3年、5年、10年”，页面写成“缴费灵活可选”	可选档位不可判，答案无法引用
责任免除	条款分项列明并含例外，页面写成“部分情况不保”	免责范围被弱化，与他页相矛盾
免责提示	条款要求显著提示并明确说明，页面小字一句带过	提示义务不成立，效力存疑[12]

比对点来自条款结构与第十七条的提示说明要求，只涉页面表述，不构成法律意见；文案仍须法务与合规审核。

逐字比对先比端点和档位，这两类可机器比对；再比免责分项，这类需人读，因为条款用法律语言、页面用语体语言。顺序反过来，人先陷进免责措辞，端点这类硬错误就被放过。

第三道约束是比较可以绝对化不行

对照页要比较，又不能下"最好""第一"这类判断。可用的写法是把判断改成位置：某款产品在某个字段上处在第几位、与中位数差多少。位置是可以复核的，判断不是。**把优劣句换成位置句，是对照页能同时满足合规与可引用的关键。**

还有一处细节：比较范围要写明。是本公司同类产品之间比，还是与全市场同类比，读者与引擎都会据此判断这句话能不能用。范围不写，位置句也会被当成绝对判断。

第四道约束是演示值与真实值分开

演示利率、演示现金价值属于销售演示材料，不是合同承诺。页面上一旦与真实条款数值同列同格式，读者会混淆。可行的做法是分栏：左栏条款数值，右栏演示示例，两栏各有小标题与口径说明，演示栏顶部标"演示口径，非合同承诺"。

这条约束有制度来源。《一年期以上人身保险产品信息披露规则》要求保险公司披露分红实现率，要求分红型产品说明书采用保证利益与红利利益两档演示，用于利益演示的利差水平分别不得高于0、4.5%减去产品预定利率，并取消高、中、低三档演示利率表述[13]。取消的是三档利率的口径，留下的是两档结构；页面上的“演示利率”不再是可自选的档位，而是有上限、有计算基础的一组固定口径。

同一规则还规定了演示旁的固定表述：利益演示须用醒目字体标明其基于精算及其他假设，不代表公司历史经营业绩，也不代表对未来经营业绩的预期，红利分配不确定；演示时不得披露用于演示的投资回报率[13]。这几句是页面用语的规定动作，缺一句，演示就滑向承诺的近义表达。

行业报道称2026年7月起分红险演示利率上限由3.9%下调至3.5%（行业观察，非官方口径）。上限一变，演示数字都要重算，页面若没有把演示与条款分开，重算就得整页重做；分开了，只动演示栏。这也是分栏写法在保险行业的实际价值：它把一次口径调整的成本，从整页重做压缩到一栏重算。

表4-3 合规约束与可引用性并不冲突的三处证据

约束越严的地方	答案越需要什么	官网的独占优势
分红不确定	一句明确的边界	只有公司能给出确定条款与不确定部分的准确分界
减保有上限	一条可核的计算规则	规则由公司设定，他人只能转述
偿付能力按季披露	一个带期次的数值	披露口径由公司按监管要求发布

本表用于说明约束与可引用性的关系，不构成对具体产品的宣传建议。

05 信任资产是可核验的披露原件

保险行业的信任资产不只在品牌声量，也在可核验的披露。偿付能力充足率按季度披露，分红实现率按年度披露，保费与赔付由监管统一统计。三类数据的共同点是：有固定口径、有期次、有发布主体，无法伪造，也无法绕过公司去获得。

表5-1 保险披露的三层可核性

披露项	发布主体	期次	当前形态与可摘取程度
行业保费与赔付	金融监管总局	月度、年度	统计表与网页，数值可直接取用[1]
全国经济与保险业总量	国家统计局	年度	公报正文，含分项与同比[2]
公司年度经营数据	上市险企年报	年度	年报摘要中的表格，需转成正文段落[3]
偿付能力与分红实现率	各公司信息披露	季度、年度	多为PDF附件，正文取不到数值[4]

形态一列描述公开页面上的呈现方式，不代表数据不公开；PDF公开可下载，但与"可被段落级引用"是两件事。

PDF为什么锁死了引用

实测某寿险公司2025年度信息披露报告，文件公开可下载，注册资本257.61亿元一类的关键数值写在报告正文里[4]。问题出在形态：数值在PDF里，站点页面没有同口径的表格。引擎抓网页时取不到这些数值，只能抓到"公开信息披露"这个入口标题。

这不是孤例。行业里多数公司的偿付能力报告、分红实现率表都以PDF或图片形式挂出，页面正文只写一句"详见附件"。结果是公开了却不可摘取：数据在，段落不在。图片形态更彻底——连检索都进不去。

还有第三种失效更隐蔽：期次过期。页面挂的是两年前的报告，链接能打开，数值也在，但答案里引用它就是引用过期数据。这类问题在披露页尤其多，因为披露文件按季归档，页面往往只挂当期链接，旧期次被覆盖但没有标注"本期已更新"。

把披露变成可摘取段落的三步

1. **表格化** 把报告里的关键表搬到页面上，一行一期，保留原始单位与小数位，不改口径。
2. **期次随文** 每个数值旁写明报告期，例如"2025年第4季度"，不写"当前"这类无法核对的时点。
3. **回链原件** 页面下方挂PDF或披露报告链接，让读者能回到原件复核；链接失效时页面数值仍在，不影响可读。

三步做完，同一份数据有了两种形态：给人看的PDF、给机器读的表格。前者承担法律效力，后者承担引用功能。二者不冲突，也不必重做披露流程。

表5-2 同一份数据的三种形态与可得性

形态	读者能否读到	引擎能否取到数值	改造动作
PDF附件	可读 需下载	难取 数值埋在文件里	页面同步一份表格
图片扫描	勉强 需放大	不可取 无文本层	改为可选中文本
网页表格	可读 直接可见	可取 一行一期	维持并按期更新
正文段落	可读 但有歧义	部分 数值与主语易脱钩	补主语与期次

可得性按"能否取到带主语的数值"判断，不按文件是否公开；四种形态可并存，网页表格不替代PDF的法律效力。

法定应披露六项，页面侧能摘到几项

《人身保险产品信息披露管理办法》第十一条列明应披露的产品信息：产品目录、产品条款、产品费率表、一年期以上产品的现金价值全表、一年期以上产品的说明书，以及监管规定的其他材料[11]。这是一份披露义务清单，不是页面形态清单。义务履行了，不等于内容可摘取。

把两者分开看，披露问题就从“有没有公开”变成“公开在哪、以什么形态公开”。同一份条款，可下载的PDF与页面上的参数表，法律效力相同，可引用性完全不同。

表5-3 六项法定披露与页面侧可摘取性（2026-11-20实测）

法定披露项	页面侧现状	可摘取	说明
产品目录	产品中心设8个分类，分类页无产品列表	不可摘取	分类导航不等于产品目录，缺逐款条目
产品条款	清单页73个链接中，条款文件0个	不可摘取	条款未在页面侧挂出，或挂出后不可定位
产品费率表	页面无费率表格	不可摘取	费率未以行列表格呈现
现金价值全表	页面无全表，仅见示例值	不可摘取	示例与全表是两件材料，不可互替
产品说明书	页面无可检索的说明书文本	不可摘取	说明书多以PDF随单发送，未上公开页
其他材料	按监管另行规定	未测	本白皮书不逐项判定

本表只判公开页面侧的可摘取性，不判断披露义务是否履行；披露亦可经产品目录、纸质说明书、APP与柜面完成，页面缺项不等于未披露。

这张表说明“官网查不到”与“公司没披露”可以是两件事：义务侧材料齐备，问题在承载形态——条款在PDF、费率在计划书、说明书随单发送、现金价值只给示例。四类都合规，四类都不可摘。

着力点不是补材料，是补承载。**同一份材料的第二种形态，成本远低于重做一份材料。**把已有表格搬到页面、把已有章节拆成段落、把已有附件挂成可回链入口，三步都不新增披露内容。

分红实现率的可核难点

分红实现率是保险披露里容易被误引用的一项。它的定义写得其实很具体：现金红利实现率等于实际派发的现金红利金额除以红利利益演示的现金红利金额，增额红利实现率与终了红利实现率同法，计算时以4.5%减去产品预定利率为利益演示基础，公司应在每年分红方案宣告后15个工作日内在官网披露[13]。定义清楚，难点落在落地：同一产品的不同账户要分别披露，增额方式还要分列终了红利，外部引用时往往只取一个百分数，省掉产品、账户与分红期间。

由此得到一条可核要求：披露页要能回答三个问题——哪一个产品、哪一个分红期间、哪一种红利方式。三个问题里有一个答不出，实现率就只是一个被反复转述的百分数。公司自己发布有天然优势：分母怎么算、哪些年度已过完整保单年度，只有它知道。把这两点写清，实现率才成为可核数据。

分级为本白皮书自定的经验口径，参照保险业披露惯例与信源可核性，非任何机构的官方分级标准。

表5-4 本白皮书的信源分级（经验口径）

级别	来源类型	可否支撑结论	使用限制
L1	监管统计、公司法定披露、监管规章与法律文本	可支撑	统计与披露须标期次与发布主体，法规须标名称与条号
L2	上市险企年报、评级机构跟踪报告	可支撑	须标报告类型与日期
L3	经纪与测评内容、行业媒体	仅作旁证	不进来源列表，只作行业观察

级别	来源类型	可否支撑结论	使用限制
L4	匿名转述、无出处榜单	不可用	不作为任何结论依据

06 从一款产品铺到一条产品线

改造的顺序比改造的规模重要。多数公司的第一反应是把在售产品全建页，页数一多，模板的缺陷会被复制到每一页上。更稳的做法是先做三款，验证模板在真实抓取下能不能被读出字段，再复制到其余产品。

图6-1 模板验证与规模复制的先后关系



款数为可数交付物口径，非排期承诺；实际铺开数量由在售产品与披露范围决定。

第三步容易被跳过。页面做出来、能打开、看着也整齐，就算完成了——但抓取脚本读不出字段，或读出的字段与页面显示不一致，这一页对引擎而言仍然等于零。验证方式很简单：静态抓取取一次正文与结构化块，确认每个字段都有独立文本节点，且结构化声明里的字段名与页面对得上。

这一步的成本极低，收益极高。3款产品验证一次，经验工时约两小时；跳过它，20款产品铺开后发现问题，返工要按产品数乘。实测某寿险公司清单页的教训就在这：一个模板承接全部分类，页面看着整齐，73个链接里没有一个是产品页[6]。

模板参数化是根因，不是界面问题

产品中心用一套参数化模板（ProductCenter_1200.html?id=栏目ID）承接八个分类，页面靠栏目ID切换内容。这种结构适合栏目导航，不适合产品详情：一个模板承载多个分类，产品只能在分类页里以列表形式出现，无法拥有独立地址。独立地址是引用的前提——没有独立地址，就没有可回链的原文。

连带的问题有三个。页面标题为空，引擎无法判断这一页是什么；canonical未设置，同一内容可能被多个地址重复收录；结构化声明缺失，产品类型无从声明。三项都源于同一个设计：模板以栏目为核心，不以产品为核心。

改造不必推翻CMS。可行路径是保留分类页做导航，新增一套单品页模板，由同一后台按产品ID生成独立地址。分类页只负责导语与产品卡片，参数交给单品页。这套改法不碰既有栏目结构，前端改动集中在新增模板与页面头部信息。

三处资源的现实约束

落地卡住的往往不是技术。条款字段要经法务确认，页面模板要IT排期，产品清单要业务部门给。三处各有节奏，串起来才是工期。可行的排法是先做不需要审批的部分：页面标题、描述、canonical三项属于技术规范，可直接改；结构化声明只要字段取自现有条款，也不需新增审批；需要审批的是新增参数页，因为涉及对外披露口径。

这个排法还有一个附带好处：前三项改完就能测出基线变化，用实测数据说服法务与业务，比用方案说服更省力。

表6-1 三类动作的依赖与阻塞点

动作	依赖方	可先行部分	阻塞点
页面头部信息	技术	标题、描述、canonical	无，可直接改

动作	依赖方	可先行部分	阻塞点
结构化声明	技术 + 产品	字段名与类型	取值须条款确认
参数页	法务 + 产品	模板与排布	对外披露口径审批
规则页	法务 + 合规	结构三句式	规则原文摘录范围
披露表格	财务 + 技术	历史期次整理	当期数据发布时点

依赖方指口径归属职能，不含组织架构差异；阻塞点仅指流程节点，不含人力排期。

表6-2 落地清单（可逐项勾选）

序	动作	可数交付物	完成
1	选定3款在售产品，逐项取条款字段	字段表3份	☐
2	建单品页模板，字段一页一排布	模板1套	☐
3	补页面标题、描述与canonical	页面6个（含3分类页）	☐
4	加产品类型结构化声明，字段与页面一致	声明模板1份	☐
5	静态抓取验证字段可读	验证记录1份	☐
6	建规则页3个（减保、分红实现率、偿付能力）	页面3个	☐
7	建对照页2个（同类三至五款并排）	对照表2张	☐
8	把披露数值表格化，标注期次	披露表2张	☐
9	模板定型后复制到在售主力产品	单品页20个	☐
10	按问句清单逐条复查覆盖	覆盖率表1份	☐

清单中的页面与表格数量为可数交付物，不含人力与工期估算。

改造会卡在哪六处

落地失败很少因为方案不对，多因为在六个具体位置上卡住。这六处都能在第02章的实测读数里找到对应：模板以栏目为核心、字段缺边界条件、口径漂移、披露仍是PDF、更新节奏失配、声明与页面不一致。它们不是执行不力，是结构问题，改一处就能解一类。六类里有三类可以在不新增审批的前提下先解：模板新增、字段补边界、声明与页面同源；另外三类涉口径与流程，须法务与财务参与。把这个次序讲清楚，比讲一个完整的改造方案更容易推动第一步。

表6-3 六类失败模式的识别与处置

失败模式	识别信号	成因	处置
模板以栏目为核心	单品打不开独立地址，产品只能进列表	后台按栏目ID生成页面	新增单品模板，分类页只做导航
字段缺边界条件	字段有值，读者仍问“含不含”	取值时省略括注与例外	回条款补端点，按表3-3六属性定稿
口径漂移	页面数值与条款不一致	两处分别维护，无单一来源	字段单一来源，页面标注条款版本
披露仍是PDF	页面只有“详见附件”一句	披露流程与页面流程分离	页面同步一张期次表，回链原件
更新节奏失配	同一页出现不同期次的数据	把多类材料塞进一页	按材料分页，各自按节奏更新
声明与页面不一致	结构化字段与正文数值不符	声明由另一批人另行维护	声明与页面同源生成，随页面发布

六类来自第02章实测与第04章约束的归纳，不覆盖组织、预算与人力因素；处置动作与表6-2的落地清单对应。

07 度量只需三个可复算的数

保险行业的GEO度量不需要复杂模型。三个数就够，且都能被别人复算：可摘取单元数、问句覆盖率、引用归因条数。排名不进这三个数，原因是排名随时间、地域、提问措辞变化，同一周内同一引擎两次提问也可能给出不同结果，无法作为结算依据。

表7-1 三个指标的算法与当前基线

指标	算法	采集方式	实测基线
可摘取单元数	参数表行数＋对照表列数＋规则页问答条数，按页累加	人工计数，同一口径复算	0
问句覆盖率	（有回答＋0.5×有但不合格）÷知识单元总数	按问句清单逐条对表	6.7%
引用归因条数	答案来源含自有域名的条数	固定问句集、固定引擎清单	待实测

前两项基线取自本白皮书第02章与本章实测样本；引用归因需固定问句集与引擎清单后采样才有比较意义，故记"待实测"，不填估算值。

> **问句覆盖率 = （有回答的单元数＋0.5×有回答但不合格的单元数）÷知识单元总数**
>
> 样本：15个知识单元，其中2个判为"有但不合格"（公司概况页、公开信息披露页可部分作答，但页面无分节、无表格，数值在PDF），13个判为"无"。代入得（0＋0.5×2）÷15＝6.7%。
> 合格线取70%，且决策层单元须全部有回答；该阈值是本白皮书自定的经验口径，用于判断知识资产是否够用，不是行业标准。

15个知识单元的构成本身也要可核，否则覆盖率是空算。本白皮书的样本按四类分配：选型类6个（增额终身寿哪个好、预算三百万怎么配、储蓄型还是保障型）、规则类5个（减保怎么算、分红实现率怎么看、退保损失多少、等待期多久、犹豫期退保怎么算）、验证类3个（公司规模多大、理赔流程怎样、偿付能力够不够）、时点类1个（2026年演示利率上限是多少）。

四类比重大致与公开检索可见面吻合，选型与规则占多数，验证类次之，时点类只有1个，波动却大。样本固定下来之后，每次复测都用同一批问句，只替换时点类里的年份。

表7-2 15个知识单元的构成与判定

类别	单元数	问句示例	主答材料	当前判定
选型类	6	增额终身寿哪个好、预算三百万怎么配	材料一＋材料二	6个无
规则类	5	减保怎么算、分红实现率怎么看、退保损失多少	材料一	5个无
验证类	3	公司规模多大、理赔流程怎样、偿付能力够不够	材料三	2个部分，1个无
时点类	1	2026年演示利率上限是多少	材料四	1个无
合计	**15**	—	—	**13无 / 2部分 / 0合格**

"部分"指页面可给出间接线索但不能完整作答；"合格"指页面有分节、有表格、数值可直接取用。判定由本白皮书逐条作出，不做多人复核。

样本固定下来，复测才有比较意义

覆盖率是一次快照，不是一条曲线。要让两次读数可比，样本、引擎、口径、判定人四项都得固定；任一项变动，前后两个数就不能相减。四步不难，难在坚持，因为每一步都有更省事的替代做法。

表7-3 复测的四步与常见走样

步骤	固定对象与产物	周期	常见走样
固定问句集	15个知识单元清单，版本化	改版记入附录C	中途增删问句，基线断裂
固定引擎清单	引擎与入口列表	季度复核	换引擎后仍与旧数据直接比较
固定抓取口径	静态抓取脚本与字段清单	口径变更须记录	改为手工复制正文，长度不可比
固定判定人	判定人与判定记录	同一人连续两季	换人后判定标准随人变化

四步对齐第07章的三个指标；问句集与引擎清单的变更须在附录C追加行，不与上一版直接比较。

还有一处更隐蔽的走样：判定标准。同一句“页面可部分作答”，不同人会给出不同结论。处置办法是先定示例再计数：把一百字以内、有主语、有期次算作“合格”，把只有标题或只有附件链接算作“无”，中间态写成一个具体例子。示例定完，两个人数的结果才会趋同。

表7-4 三类误差来源与处置

误差来源	表现	处置
引擎侧波动	同一问句两次答案来源不同	同一问句取三次，记录来源分布，不以单次为准
页面侧改版	读数突变，内容其实未变	抓取前后记录页面指纹，按指纹判断是否改版
人工计数口径	不同人对“一行”“一条”的计法不同	先定计数示例，再计数；示例随口径版本存档

三类误差不改变指标定义，只改变采样次数与记录方式；误差范围写进复测记录，不写进结论。

这也是为什么收敛按季度算、不按月算。月度读数受引擎波动与语料更新影响，单月差值不足以判断改造是否有效；季度跨度里，页面改版、条款更新、披露换期都会发生，三个数才看得出趋势。按季收敛不等于按季才动手，页面改造可以按月推进，读数按季比较。

为什么排名不能进结算

排名的问题有三层。第一层是采样：同一问句在不同时间问，答案来源可能不同，缓存与语料更新都会影响结果。第二层是措辞：用户问"增额寿哪个好"与"增额终身寿险哪家强"，命中的内容可能不是同一批。第三层是因果：答案里出现某个来源，不等于这个来源的页面做得好，也可能只是这一批语料里只有它写全了。

三层叠起来，排名可以作观察，不能作验收。可结算的是可摘取单元数这类静态可数的量，它不随提问措辞变化，任何人打开页面都能自己数一遍。**能自己数一遍的量，才适合写进合同。**

监测节奏与失效信号

按月复测，用同一套问句、同一批引擎。出现三种情况即视为失效，需要回查页面而不是改问句：一是页面字段改了但引用数下降；二是同一问句在两次复测中答案来源完全换了一批；三是引用数上升但落地页承接不住，用户到站后找不到对应产品。

第三种容易被忽略。引用增加说明页面被取到了，承接不住说明取到的东西与用户预期不符——可能是答案里引的是规则页，用户点进去想买产品却找不到购买入口。这类失效不改页面内容，改页面之间的导航关系。

还有一项要一起看：条款更新后的同步率。新产品上线、旧产品停售、条款修订，三类事件发生后，页面字段要在同一周期内更新。同步率不需要精确到天，按月对齐就够；漏更新一个字段，就等于在答案里留了一处过期数据。

08 本白皮书没有回答的问题

把没答的说清楚，比把答过的话再说一遍有用。以下四项本白皮书没有数据支撑，不做判断。

- **各引擎对保险类问句的召回权重**　引擎未公开排序权重，本白皮书只测页面可读性，无法推出某类页面在不同引擎里的相对位置。
- **分红实现率的行业分布**　需逐家逐产品抽样，样本不足时会得出以偏概全的结论，本白皮书不给定性描述。
- **短视频语料对答案的影响**　保险类问句的语料里短视频占比不低，但视频内容无法直接抓取文本，影响程度未测。
- **APP与小程序的产品页表现**　本白皮书只测公开网页，APP内页面与外链跳转的抓取差异未纳入。

表8-1　四项未答问题各自影响哪一章

未答问题	影响的章节	若结论相反会怎样	补证方式
引擎召回权重	第三章	三类页面的优先次序需要重排	多引擎对照采样
分红实现率分布	第五章	披露资产的价值被高估或低估	逐产品抽样
短视频语料影响	第七章	文本页面的作用被稀释	语料结构统计
APP页面表现	第六章	网页端改造收益被分流	渠道承接比例测算

"若结论相反"一列只给方向，不给幅度；补证方式为方法建议，本白皮书未实施。

四项里有两项（视频语料、APP页面）会显著改变结论的适用范围。如果一家公司的产品主要在APP内成交，网页端改造的收益会被分流；如果其目标问句集中在视频平台，文本页面的作用会被稀释。这两类公司做改造前，应先测一次自有渠道的问句承接比例，再决定投入重心。

本白皮书的判断，压缩成三句

保险行业在生成式引擎里的位置，不取决于写了多少文章，取决于条款被转译成字段的比例。规则页是别的机构抄不走的资产，也是容易被写成营销稿而报废的资产。度量用三个可复算的数，不用排名。

这三句里没有一句是"多做内容"。反过来说，如果一家寿险公司明年只做一件事，把在售主力产品的条款字段化成参数页，比做一百篇行业观察更有用——前者产出的是别人可以引用的段落，后者产出的是又一篇别人可以改写的文本。

附录A 术语与口径对照

表A-1　术语与口径对照

术语	本白皮书的口径	易混用法与区别
可摘取单元	能被单独取走并独立成立的一小段，含参数表的一行、对照表的一列、规则页的一条	不等于段落。一段议论不可摘，一张表的某行可摘
知识单元	对应一个真实问句的基本回答单位，本白皮书取15个	不等于页面。一页可承载多个知识单元
结构化声明	页面内以JSON-LD形式声明内容类型与字段	不等于标签优化。声明要能被解析且与页面一致

术语	本白皮书的口径	易混用法与区别
独立地址	单款产品拥有的可回链页面地址	不等于分类页参数。带参数的分类页不构成独立地址
披露原件	由公司按监管要求发布、有期次与发布主体的数据文件	不等于新闻稿。原件有法律效力，转述没有
演示口径	销售演示材料中的假设与算法，非合同承诺	不等于条款口径。演示值可变，条款值固定
规则页	解释一条由公司设定的规则及其条款依据的页面	不等于产品页。规则页不讲产品优势
法定披露	按监管规章应公开的材料：目录、条款、费率表、现金价值全表、说明书	不等于可摘取。披露合规但形态为PDF时，答案取不到内容
可摘取性	页面内容能否被单独取走、带主语与期次独立成立	不等于公开性。公开可见的内容未必可摘

本表口径仅适用于本白皮书，与第11册的五级考核口径不是同一套定义。

附录B 一款产品的字段化样张

字段名为本白皮书建议的基础集，可增列，不得为版面整齐删减限定条件；括号内为需按备案信息与条款原文填入的内容，不预填、不估算。

表B-1 产品字段化样张的字段与取值来源

字段	取值来源	样张占位
产品名称	备案名原文	（备案名全称，不缩写）
险种类型	条款首段	（终身寿险 / 年金保险 / 医疗保险等）
投保年龄	条款投保条件	（起止年龄，保留括注与是否含端点）
缴费期	条款缴费条款	（档位原文，如3年、5年、10年）
保额规则	条款保险金额	（计算方式原文，含系数与基数）
减保规则	条款保单权益	（年度上限、起取时间、计算基数）
等待期	条款责任免除	（期数原文；无等待期的写明"无"）
现金价值示例	费率表	（示例口径与假设条件须同时写出）
分红说明	条款与产品说明	（分红不确定的原文表述，不改写）
适用人群	条款与产品说明	（按条款与说明原文概括，不加倾向性评价）
条款文件	监管备案文件	（文件名称与备案编号）
页面更新日	站点发布记录	（年月日，便于读者判断时效）

附录C 版本与修订记录

后续版本在此表追加行；口径变更须在同行注明变更点，不与上一版直接比较。

表C-1 版本与修订记录

版本	日期	修订内容	数据口径
V1.0	2026年11月	首版：答案供给结构、分险种结构数据、四家官网实测、产品链实测、三类页面与字段六属性、四道约束与法规依据、法定披露与可摘取性、落地失败模式、三指标度量与复测口径	官网抓取2026-11-20；行业与公司数据取2025年度

参考来源

十三条来源按用途分三类：法规与规章用于界定字段与披露义务，统计与公司披露用于取数，实测记录用于页面读数。技术规范只用于解释解析与爬虫行为，不用于行业判断。

[1] 国家金融监督管理总局：《2025年12月人身险公司经营情况表》，统计与风险监测司，2026-01-30
[2] 国家统计局：《中华人民共和国2025年国民经济和社会发展统计公报》，2026-02-28
[3] 中国人民保险集团股份有限公司：《2025年年度报告摘要》，上海证券报，2026-03-27
[4] 中国人民人寿保险股份有限公司：《2025年度信息披露报告》，2026-04-30
[5] 中债资信：《中国人民保险集团股份有限公司2023年资本补充债券2026年定期跟踪评级》，中国金融信息网，2026-06-30
[6] 本白皮书实测：四家寿险公司官网首页及某寿险公司产品入口页的静态抓取，2026-11-20
[7] IETF：《RFC 9309 Robots Exclusion Protocol》，2022-09
[8] W3C：《JSON-LD 1.1 Processing Algorithms and API》，2020-07
[9] 中国人寿保险股份有限公司官网首页，静态抓取对照站点，2026-11-20
[10] 中国太平洋人寿保险股份有限公司官网首页，静态抓取对照站点，2026-11-20
[11] 中国银行保险监督管理委员会：《人身保险产品信息披露管理办法》（银保监会令2022年第8号），2022-11-11
[12] 全国人民代表大会常务委员会：《中华人民共和国保险法》（2015年修正），2015-04-24
[13] 中国银行保险监督管理委员会：《一年期以上人身保险产品信息披露规则》，2023-01

引用与核验说明

表D-1　来源编号与正文位置索引

编号	来源	支撑的章节	级别	形态
[1]	金融监管总局人身险经营情况表	第一章、第二章	L1	网页统计表
[2]	国家统计局国民经济和社会发展统计公报	第一章	L1	网页正文
[3]	中国人保2025年年度报告摘要	第二章	L2	报纸版面
[4]	人保寿险2025年度信息披露报告	第五章	L1	PDF
[5]	中债资信跟踪评级报告	第二章	L2	网页
[6]	本白皮书站点实测	第二章、第七章	自有	实测记录
[7]	RFC9309爬虫排除协议	第二章	规范	技术文档
[8]	W3C JSON-LD规范	第三章	规范	技术文档
[9]	中国人寿官网首页	第二章	实测	对照站点
[10]	太保寿险官网首页	第二章	实测	对照站点
[11]	人身保险产品信息披露管理办法	第三章、第五章	L1	规章文本
[12]	中华人民共和国保险法	第四章	L1	法律文本
[13]	一年期以上人身保险产品信息披露规则	第四章	L1	规章文本

级别按表5-4口径标注；"自有"为实测记录，"规范"为技术文档，"实测"为对照抓取。

十三条来源里，十二条有可直接访问的公开地址：十条为监管统计、公司披露、对照站点与法规规章文本，两条为技术规范文档；一条为自有实测记录，无外部地址。编号与正文的对应是单向的：正文每处方括号编号只指向一条来源，来源列表每条至少被正文引用一次。分红险演示利率上限由3.9%下调至3.5%这一具体数值来自行业报道，未列入编号，正文已注明其性质；该项的制度依据列入[13]。

实测数据的核验方式是可复算：同一页面、同一口径、同一字段清单，任何人重做一次应得到同一组读数；差异通常来自页面改版，届时以新版读数为准并在附录C追加修订行。

复算清单

五步产物为字段表、覆盖率表、对照表与复测记录，可独立完成，均不含排名。

1. **定问句** 三类各取一个（选型、规则、验证），记录引擎、时间与完整问句。
2. **抓页面** 静态抓取目标页，记录HTTP状态、正文长度、标题、canonical与结构化块数，脚本同版本存档。
3. **判字段** 按表3-2十项逐条判定，缺失记0，有值但无边界条件记0.5。
4. **算覆盖** 按（有回答＋0.5×部分作答）÷15计覆盖率，合格线70%，注明各部分判定依据。
5. **留指纹** 记录页面指纹与引擎入口，与上一版并列存放，差异逐条注明。

五个步骤产出四份材料，每份材料都有固定字段与归档方式，缺一份，读数就缺一段依据。

表D-2 复算产物的四份材料

产物	内容	用途	归档方式
字段表	十项字段的判定、依据与缺失项	定位缺哪一类字段	与页面指纹同期存档
覆盖率表	15个单元的判定与算式	判断知识资产是否够用	按季度留存
对照表	同类产品的并列字段	支撑比较型问句	随产品变更修订
复测记录	引擎、问句、指纹、判定人	保证前后两次可比	每轮一档，不合并

四份材料均不含排名；归档要求与附录C的修订行对应，改版与口径变更都在修订行里留痕。

四份材料与第07章的三项指标一一对应：字段表对应可摘取单元数，覆盖率表对应问句覆盖率，复测记录承载引用归因条数的取样条件。三项指标都要能被第三方按同一份材料复算；材料不全，指标就退回成一个无法核对的数字。

判定示例要写死三档，先定示例再计数，否则换个人数出来的结果会不同：

- **合格** 页面有分节标题、有表格或分行字段，数值带单位与期次，如“2025年第4季度综合偿付能力充足率”后接具体百分比。
- **部分** 页面只有标题或只有附件入口，正文无对应数值，如“公开信息披露”入口页。
- **无** 页面不含相关主体的任何数值或规则描述。

三档示例一经确定就写入复测记录，后续各轮沿用；需要调整时单列一行说明，不默默改口径。这也是第07章那条判定的落地写法：示例先于计数，口径先于结论。

本白皮书不含对任何具体产品的收益预测、销售建议与购买引导。文中出现的产品名称仅用于说明页面结构问题，不构成推荐。企业按其自身情况做改造决策时，条款与宣传文案须经内部法务与合规审核，本白皮书不替代该流程。