Sánchez、Feijóo还是Abascal？西班牙29N大选三大政党领导人X帖子中的发言者归属与意识形态一致性：大语言模型角色研究

Víctor Yeste[1,2,*], Candida Maria Greco[1,*], Paolo Rosso[1,3],Andrea Tagarelli[4]

[1]瓦伦西亚理工大学PRHLT研究中心，西班牙[2]瓦伦西亚欧洲大学科学、工程与设计学院，西班牙[3]瓦伦西亚人工智能研究生院与研究网络（ValgrAI）[4]卡拉布里亚大学DIMES系，伦德，意大利 {vicyesmo, candida.greco}@upv.es, prosso@dsic.upv.es, tagarelli@dimes.unical.it

arXiv:2611.29765v2 [cs.CL] 24 Nov 2026

摘要

来源线索是政治传播的核心，但以大语言模型（LLM）角色作为合成受众的接受研究，尚未检验这些角色是否会对发言者是谁作出反应，而不只是关注说了什么。考虑到西班牙将于2026年11月29日举行大选，我们开展了一项发言者归属实验，使用了西班牙工人社会党、人民党和呼声党领导人的1,075条匿名化X帖子，涉及四个主题（经济与就业、家庭开支、移民和腐败），以及791个基于价值观、由大语言模型生成的不同人物画像。每条帖子分配给300个按主题选取的角色档案，其中150个以Inglehart–Welzel文化地图为依据，另外150个以道德基础理论为依据，并设定四种条件：无作者、Sánchez、Feijóo或Abascal，由此得到322,470个完整的角色—帖子四元组。无论标注哪位领导人，都会降低受众对相同文本的支持度；Abascal带来的惩罚几乎是Sánchez和Feijóo的两倍。这种惩罚取决于角色在引导下表达的左–右自我定位（意识形态一致性）：左派角色会对Sánchez网开一面，右翼角色则会对Feijóo网开一面，尽管消息评估提示词中从未提及意识形态。无论标注谁，腐败帖子都会受到惩罚；而在其他主题上，意识形态上最接近的领导人通常几乎不会带来惩罚，甚至完全没有惩罚。领导人排序与48期CIS晴雨表（2022–2026年）中三个群体的公众舆论及全部12项差异的方向一致，但与模型2024年知识截止日期之前的晴雨表相比，对近期晴雨表的拟合较差。我们的研究结果表明，大语言模型角色会通过系统性的、与价值观一致的先验来评价政治发言者：然而，尽管它们在调查晴雨表中复现了人类对候选人的排序，这种对齐只是序数层面的，而非 quan

数量化，凸显了它们在受控实验中的实用性作为信息来源效应实验的工具——但不能替代实证公众舆论调查。

1 引言

在政治传播中，说话者是信息的一部分。同一项提议若出自本党而非竞争对手，受到的评价就会不同；政党线索甚至可能比政策Cohen，2003；Druckman 等人，2013）的内容更有影响力。如今，大语言模型已被用于模拟政治受众，以人物画像代表不同公民群体。据我们所知，**这些人物画像是否会对谁在发言作出回应，而不只是对说了什么作出回应，尚未经过检验**，因为采用合成受众的接受研究，要么始终显示消息作者，要么始终将其隐藏。我们拓展了 Greco 等人（2026b）的研究；该研究考察了价值导向人物画像如何接收一位身份固定且经过匿名化的行动者所发的帖子。由于说话者身份没有变化，该设计无法将来源线索效应与信息内容区分开来。本研究在合成受众中开展了经典的来源线索实验。我们选取了西班牙三大政党领导人佩德罗·桑切斯（西班牙工人社会党）、阿尔贝托·努涅斯·费霍（人民党）和圣地亚哥·阿巴斯卡尔（呼声党）发布的 1,075 条匿名化 X帖子，主题涉及四个重要议题：经济与就业、家庭开支、移民和腐败。实验使用 gpt-oss-20b 实例化了 791 个不同的价值导向人物画像；针对每个主题，其中 150 份角色档案以英格尔哈特–韦尔策尔文化地图（IW；Inglehart 和 Welzel，2005）为基础，另有 150 份以道德基础理论（MFT；Graham 等人，2013）为基础，对每条帖子均在四种条件下进行评价，条件之间唯一的差异是显示的作者：无作者、Sánchez、Feijóo 或 Abascal。由于每条帖子都会以每个名字显示，因此可以在同一受众中，针对完全相同的文本测量名字的影响

[*]贡献相同。

群体。西班牙非常适合探讨这一问题。其政党体系包括中左翼政府、中右翼主要反对党和激进右翼第三党，因此可以将作者归因与意识形态距离交叉考察。西班牙社会学研究中心（Centro de Investigaciones Sociológicas，CIS）每月发布晴雨表，对三位领导人进行评价，并询问左右自我定位问题；我们也向人物画像提出了同样的问题。

我们提出四个问题。**研究问题1**：将同一条消息归因于 Sánchez、Feijóo 或 Abascal，而不是匿名呈现，会如何改变对它的评价？**研究问题2**：这种效应是否取决于引出的左–右自我定位，因此人物画像会对与自身立场最接近的领导人作出更积极的回应？**RQ3**：归因效应的大小或方向是否会因显著政治议题而异？**RQ4**：左派、中间派和右翼人物画像对三位领导人的排序，是否与相应的 CIS 受访者群体相同？

点名任何一位领导人都会降低对同一条消息的支持度，而点名 Abascal 所造成的降幅约为另外两位的两倍（在从 -1 到 $+1$ 的立场量表上，分别为 -0.10、-0.10 和 -0.19）。总体效应掩盖了截然相反的反应：左翼人物画像对 Sánchez 的惩罚最轻、对 Abascal 的惩罚最重；右翼人物画像对 Feijóo 的惩罚最轻、对 Sánchez 的惩罚最重；三项一致性对比均为正值，尽管消息评估提示词从未包含人物画像引出的意识形态定位。只有在腐败议题上，每个群体对所有名字都会降低支持度。在其他议题上，最接近的领导人之名不会带来惩罚。各群体偏好三位领导人的排序，与 2022 年至 2026 年间 48 期 CIS 民意调查中发现的排序一致；这种一致性体现在所有群体及全部十二项差异的方向上，但这种吻合有时间上的局限：在模型知识截止日期为 2024 年 6 月之前的每一期民意调查中均成立，而从 2025 年起，这种吻合在右翼受访者中有所减弱。

我们的贡献包括：（1）首次为合成受众设计完全交叉的发言者归因方案，结合匿名化和反事实归因，并通过实验前审计检验模型是否能识别领导人；（2）通过引导方式 elicited 的 CIS 式自我定位，其与人物画像底层道德和文化价值观的关联，在方向上与人类样本中观察到的关系一致；（3）与 48 项 CIS 民意调查进行比较，仅限于顺序和方向，同时也为人物画像所呈现的领导人形象标注了时间；以及（4）发布代码、配置、审计工件、帖子标识符和派生标签，但不发布帖子文本。[1]

2 相关工作

大语言模型角色作为合成受众。以人口统计或传记概况为条件的大语言模型，已被用于模拟调查受访者（Argyle et al., 2023），并在接受调查数据微调后，预测各子人口如何分配其回答（Suh et al., 2025）。同一领域的文献也记录了这种方法的局限：模型反映的是不均衡的默认倾向，而非被要求扮演的群体的观点（Santurkar et al., 2023）；对于那些特征很少共同出现的概况，模型更难予以引导（Liu et al., 2024）；模型会扁平化其所呈现的群体（Wang et al., 2025）；与具有指导性的提示相比，模型同样会从无关或误导性的提示中学习（Webson and Pavlick, 2022）；并且会随着人格提示形式的变化而改变其模拟内容（Lutz et al., 2025; Weeber et al., 2026）。合成样本可以再现平均值，却无法捕捉真实意见的协方差结构，并且会加入自身的偏差（Bisbee et al., 2024; Boelaert et al., 2025; von der Heyde et al., 2026）；一项对63项人格研究的综述呼吁使用以经验数据为基础的人格，并明确说明目标人口（Batzner et al., 2025）。近期有两项研究与我们的设置较为接近。左倾与右倾的人格会使从真实媒体中提取的说服性内容的支持度产生两极分化（Antelmi et al., 2026）；政治人物画像会根据证据与其指定身份的契合程度来评判证据（Dash et al., 2026）；但两项研究都没有改变说话者。我们对人格的使用范围比调查模拟更窄。本文中的每一项比较，都是在相同受众评判的条件之间进行的；而与调查数据的比较仅涉及顺序和方向。

政治传播中的来源线索。信息传递者会改变信息的接受方式，这是政治意见研究中证据最确凿的发现之一。同一项政策在贴上政党标签后会受到不同评价（Cohen，2003）；来自对立政党的线索对评价的影响甚至大于来自本党派的线索（Nicholson，2012）；精英极化越严重，这种效应越明显（Druck- man 等人，2013），而且这种效应可能凌驾于受访者自述的意识形态之上（Barber 和 Pope，2019）；不过，掌握政策信息的公民对政策信息的权重至少不亚于对来源线索的权重（Bullock，2011）。

[1] https://github.com/VictorMYeste/ 西班牙语-政治-传播

西班牙是一个颇具代表性的研究案例，其政党体系的情感极化程度位居欧洲最高之列（Reiljan，2020；Torcal 和 Comellas，2022）。接触领导人本人的账号则是另一回事：在2019年欧洲议会竞选期间开展的一项面板准实验中，关注候选人的 Twitter 账号并未加剧情感极化（Lorenzo-Rodríguez 和 Torcal，2022）。语言模型对这些领导人已有既定看法：在14种模型中，对 Sánchez 的情感倾向为中立至正面，对 Feijóo 的评价较为温和，而对 Abascal 的评价则明显负面（Goyanes 等人，2025）。模型也会受到说话者身份的影响。指出某条陈述的来源，会破坏对该陈述进行评分的模型之间的一致性（Germani 和 Spitale，2025）；政党提示也会改变模型对同一陈述给出的情感标签（Vallejo Vera 和 Driggers，2025）。新闻媒体的名称会将模型对一篇内容不变的文章所判定的政治偏见，拉向该媒体已知的立场（Park 和 Kim，2025）。句子中具名的政治人物会改变其情感分类；若人名是虚构的，这种影响则较弱（Elbouanani 等人，2025）。一项审计分析了19亿个数据点，其中只有实体名称不同，结果发现模型会惩罚右翼政治人物、偏袒左派政治人物（Elbouanani 等人，2026）。这些研究通常将来源导致的变化视为模型偏见或标注不稳定性。相较之下，我们将其视为受角色人格影响的接受效应：在保持信息内容不变的情况下，我们考察领导人姓名的整体效应，以及这种效应如何随价值观导向的角色档案而变化。这一设计将普遍存在的说话者归因效应，与角色价值观相关的异质性区分开来，但并未消除底层模型的政治先验（Feng 等人，2023；Rozado，2024；Motoki 等人，2024；Bernardelle 等人，2025）。

价值观框架与道德接受。 这些人物画像依据两个框架选取：道德基础理论及其测量工具道德基础问卷-2（Graham 等人，2013年；Atari 等人，2023年）以及英格尔哈特–韦尔策尔文化地图（Inglehart 和 Welzel，2005 年）。在人类样本中，两者都与政治意识形态相关；自由派和保守派对各项道德基础的重视程度有所不同（Graham 等人，2009年；Kivikangas 等人，2021年），这使我们能够检验这些人物画像被引导作出的位置是否与其价值观相符。所用的人物画像资源来自Greco 等人 (2026a)，他们基于世界价值观调查变量生成人物画像，并检验其在英格尔哈特–韦尔策尔地图上的位置及其道德档案。接受度研究方案源自 Greco 等人(2026b)，他们研究了这类人物画像如何接收单一行动者传递的消息。本文在该方案中加入了说话者操控和意识形态层。

3 研究设计

每个人物画像 p 都会在其指定主题中，针对四种发言者条件 c评估每条匿名帖子 m：无作者、Sánchez、Feijóo 或 Abascal。对于每种具名条件$c \in$ {Sánchez、Feijóo、Abascal}，说话者归因效应为 $\Delta_{p,\,m,\,c} = y_{p,\,m,\,c} - y_{p,\,m,\,\text{anon}}$。只有在四种条件下均有观测结果时，人物画像—帖子配对才会纳入分析，从而形成完整四元组。因此，每种效应都是配对内比较。

该设计采用完全交叉方式：无论帖子由谁撰写，每条帖子都会以全部三个姓名呈现。因此，姓名之间的差异不可能源于帖子，因为每条帖子都会出现在每种条件下；也不可能源于人物画像，因为每个人物画像都会在每种条件下评判相同的帖子；差异必定来自姓名本身。反事实归因只存在于提示词中：每条帖子的真实作者都已知，并用于分析，但绝不会展示给模型。

我们的研究问题从四个角度考察 Δ (图1)：研究问题1针对每个名字，计算其在全部人物画像和帖子中的平均值；研究问题2根据人物画像通过单独调用所引出的左右自我定位进行拆分，该调用不会进入判断提示词；研究问题3按主题进行拆分；研究问题4则比较三位领导人在左派、中间派和右翼人物画像中的平均说话者归因效应所形成的排名，以及其在相应意识形态群体中的CIS受访者平均评分所形成的排名。

4 数据

4.1 帖子

我们通过 X API v2 的完整档案搜索，收集了三位领导人的官方 X 账号在2022年4月1日（Feijóo 成为人民党领导人的日期）至2026年10月2日期间发布的所有帖子，并排除了转发。共收集到10,863条帖子，其中7,625条为原创帖子。[2] 仅将原创帖子用作

[2]API统计的八条帖子在下载前已被删除。语料库由仍然

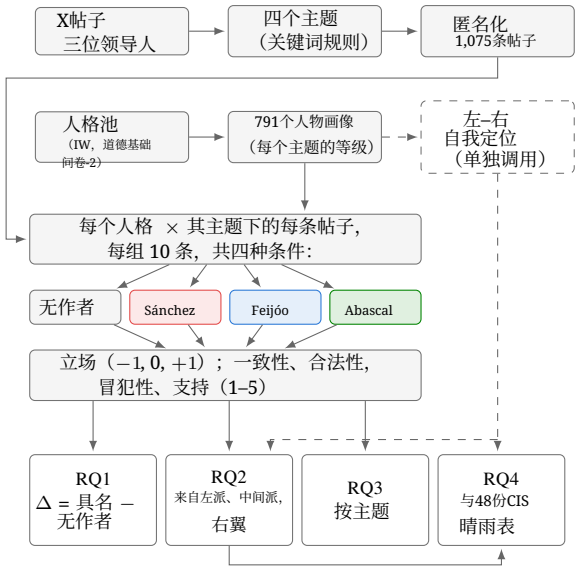

图1：研究设计。三位领导人的匿名帖子由价值导向人物画像在四种条件下进行评判，四种条件的区别仅在于显示的作者。每个人物画像的左右自我定位通过单独调用（虚线）进行 elicitation，且在其评判帖子时绝不显示；该定位用于将人物画像分组，以回答研究问题2，并将其与CIS受访者匹配，以回答RQ4。

刺激材料，因为引述和回复依赖另一位作者的文本；若将其作为刺激材料，该文本会缺失，或引入第二种往往可以辨认的声音。我们采用了 Greco 等人(2026b) 的质量筛选标准，剔除过短或信息量过低的帖子（此处指少于40个字符或7个词的帖子），并删除了重复帖文。

4.2 主题

在下载任何数据之前，我们已按规则确定主题，以使其与公众关注度以及领导人实际发布帖子的内容相对应。对于每个议题，我们计算了窗口期内有可用表格的46期CIS月度晴雨表中，受访者将其列为西班牙三大问题之一的比例的平均值（CIS的问题允许受访者最多列出三个问题），并选取显著性最高且每位领导人至少发布过十条原创帖子的议题。由此选出四个主题：(i) 经济与就业，(ii) 家庭开支（住房、能源和价格），(iii) 移民与边境，以及 (iv) 腐败。健康的显著性排名第三，但由于Abascal只发布过四条相关帖子，因此未纳入。

帖子根据每个主题的一系列西班牙语关键词和短语进行匹配，这些关键词和短语以提高精确度为目的编制（附录 A.1），未匹配任何主题或同时匹配两个主题的帖子均被剔除。在一次人工检查中

于2026年10月2日公开。

主题	Sánchez	Feijóo	Abascal	总计
经济、就业	304	99	30	433
家庭开支	123	94	11	228
移民、边境	50	40	56	146
腐败	27	127	114	268
总计	504	360	211 1	,075

表1：按主题和真实作者划分的刺激材料：通过质量与匿名化检查的四个主题上的原始 X 帖子。每条帖子均在四种条件下进行判断。

对80条匹配帖子进行检查后发现，96%的帖子在主题含义下使用了该关键词，86%的帖子主要围绕该主题展开。表1 按主题和真实作者列出了1,075条刺激材料。其中两个单元格的数量较少（家庭开支主题下的Abascal、腐败主题下的Sánchez），但由于每条帖子都以三个姓名进行判断，这只会限制按真实作者开展的分析。

4.3 匿名化与归因

在提示词输入前，三位领导人、他们所属的政党、竞选话题标签和用户账号均被替换为占位符（[政治_行动者]、[政党]、[竞选_标签]、[用户]）。其他政治人物的姓名予以保留，因为替换这些姓名会使许多帖子变得空洞，尤其是涉及腐败的帖子。包含残留姓名或变得过短的帖子均被排除，因此原有的7,625条帖子中剩下7,299条（附录A.1，表6）。四种条件呈现的匿名化文本完全相同，区别仅在于提示词中的归因行。

匿名化会隐藏姓名，但不会隐藏立场。一条代表政府发声的帖子符合 Sánchez，而不符合其他任何人。针对这类标记（代表政府发声或反政府、抨击右翼、左派或人民党）的关键词规则，将46%的刺激材料标记为至少在一个姓名下不太可能出现，尽管它漏掉的情况多于捕捉到的情况。若据此筛选，每个单元格中的帖子数量将过少，因此我们保留完全交叉设计，并分别报告真实归因、合理归因和不合理归因的效果（附录A.5，表13）。

4.4 人物画像

我们使用Greco 等人(2026a) 的人物画像资源，正如 Greco 等人(2026b) 所做的那样。这些人物画像是基于世界价值观调查的十个变量构建的大语言模型生成的档案，每份档案都包含一段简短传记、英格尔哈特–韦尔策尔（IW）坐标和道德基础问卷-2（MFQ-2）道德基础得分（Inglehart 和 Welzel，2005；Atari 等人，2023年）。与该论文一样，每

个人物画像都在两种框架下进行选择，这两种框架也决定了模型能看到的内容：IW人物画像由其传记以及生存–自我表达和传统–世俗理性坐标描述；MFT人物画像则由其传记、六项基础维度得分和36个MFQ-2题项答案描述。针对每个主题和框架，我们在三个主题条件化档案层级（高、中和低；附录B.1）中每个层级选取50个人物画像。这样，每个框架有600个名额，共有791份框架专属角色档案（441份IW和350份MFT），因为部分人物画像会用于多个主题。这些层级在价值空间中定义，且绝不会出现在提示词中（附录B.1，表15）。意识形态仅通过下文所述的自我定位体现。

该资源没有左–右变量，因此我们自行获取了这一数据。每个人物画像都在单独调用中回答了一次 CIS 自我定位问题（从 1 表示西班牙政治中最靠左，到 10 表示最靠右）；调用使用相同的系统提示和人物画像文本，且不包含帖子。我们将人物画像分为左派（1–4）、中间派（5–6）和右翼（7–10），并对 CIS 受访者采用相同的分界点。1–10 量表没有单一的中点，因此两个居中的数值构成中间派（Aybar 等人，2024年），而对 CIS 系列极化问题的研究也采用相同的左右区间（Torcal 和 Carty，2026年）。图2还分别报告了每个位置。IW人物画像中，三组分别占27%、44% 和 28%；MFT人物画像中，分别占39%、26% 和 35%。回答反映的是模型对人物画像的解读，仅用于对其分组，但这并非任意划分：尽管人物画像描述中没有明确的意识形态标签，且所获取的位置从未纳入消息评估提示，但自我定位与向模型展示的价值观相符，其方向也与人类样本中观察到的关系一致。MFT人物画像在关怀和平等方面越高，位置越偏左；在比例原则、权威和忠诚方面越高，位置越偏右（斯皮尔曼 ρ 从 -0.87 到0.67）；IW人物画像则在自我表达价值观方面越高，位置越偏左（$\rho = -0.60$）。因此，研究问题2和 RQ4 中使用的意识形态差异，与向模型展示的价值档案存在系统性关联（附录B.3）。

4.5 CIS民意调查

对于 RQ4，我们使用 2022年4月至2026年10月期间的48期 CIS月度晴雨表微观数据，该期间与这些帖子发布的时间相对应（社会学研究中心，2026b）。每期晴雨表都会询问受访者，对各位领导人进行评分（1至10分），并请他们在与人物画像相同的左右量表上为自己定位。我们保留了193,263名受访者中的175,764名，他们均为自己定位，并对三位领导人全部进行了评分；随后计算每期晴雨表内的加权平均评分，再对各期晴雨表进行等权平均（附录 D）。

要求受访者对各位领导人进行1至10分的评分，并在与人物画像相同的左右量表上为自己定位。我们保留了193,263名受访者中的175,764名，他们均为自己定位，并对三位领导人全部进行了评分；随后计算每期晴雨表内的加权平均评分，再对各期晴雨表进行等权平均（附录 D）。

5 实验设置

模型与解码。 所有判断均由 `gpt-oss-20b` (OpenAI，2025）生成。这是一款开放权重模型，声明的知识截止日期为2024年6月，也是 Greco 等人(2026b）所用的模型。模型通过 vLLM 提供服务，并经由兼容 OpenAI 的 API 调用；温度设为 0，使用固定随机种子，令牌数上限为 4,096，并以 JSON 格式输出。每次调用都会向一个人格呈现一组属于同一主题的十条帖子，并设定一种条件。生成任务使用四块 NVIDIA A30 GPU 运行，每块显存为 24 GB；每块 GPU 上均有一个独立的 vLLM 副本提供相同模型服务。经济与就业、移民与边境主题的运行于2026年10月27日完成，家庭成本与腐败主题的运行于2026年11月2日完成。

提示词。 系统提示要求模型扮演相应人格，将显示的归属信息视为既定事实；如果未具名说话者，则不得根据风格、主题或外部知识推断其身份。用户提示包含人格档案、一行说明这些消息涉及西班牙当前政治议题的文字、条件行（“以下每条消息均归属于 [姓名]”或“未提供说话者身份”）、评估规则、十条匿名化消息以及输出模式。提示词使用英语，帖子则保留西班牙语。对于每条消息，模型都会返回一个立场（支持、中立或反对）和四项 1–5 评分：一致性、合法性、冒犯性和支持（即该人格是否会原样分享这条消息），另有两个我们不作分析的简短自由文本字段（完整模板见附录 A.3，图 3）。在322,500个人设—帖子对中，有322,470对（99.99%）在四种条件下均有观测值，并构成分析集（附录 A.6，表 14）。

知识审计。 实验前，我们检查了模型是否认识这三个姓名。我们针对每位领导人和三个对照名称各询问了20次，其中两个是虚构但可信的

西班牙姓名和一个不透明代码。模型每次都能识别出各位领导人，并正确说出其所属政党；在左右量表上，Sánchez、Feijóo 和 Abascal 的得分分别为 3、8 和 9，波动不超过 1 分，而且从未识别出任何对照名称（附录 A.4，表 10）。这证实了模型对研究问题2中所用领导人的左–右排序，也表明归因行呈现了预期的身份。

估计。 立场编码为 -1、0 和 $+1$，评分仍采用 1–5 分制。我们报告Δ 在完整四元组上的平均值，以及按消息和人设进行双向聚类稳健处理后计算的 95% 置信区间（Cameron 等人，2011）。一致性通过以下对比进行检验：$(\Delta_A - \Delta_B)^{\text{left}} - (\Delta_A - \Delta_B)^{\text{right}}$；对于 A 在左、B 在右的领导人，当左翼群体相较于右翼群体对归因于 A而非 B的说话者归因反应更积极时，该对比为正。对于 CIS，我们计算每个晴雨表中的加权平均评分和领导人之间的差异，采用线性化标准误，并对各晴雨表赋予相同权重后进行合并。由于评分和说话者归因效应的量表不同，我们只比较各群体内领导人的排序以及每项差异的正负号。

6 结果

6.1 说话者归因（研究问题1）

提及三位领导人中的任何一位，都会降低相较于同一消息匿名版本的支持度（表 2）。Sánchez 和 Feijóo 产生的效应几乎完全相同（$\Delta = -0.100$ 两者皆如此），而 Abascal 带来的惩罚几乎是其两倍（$\Delta = -0.191$）；三者的置信区间均不包含零。归因后，大多数回答的类别没有改变（约为 65–70%）。不过，变化并不对称：回答变得较不支持的情况多于变得较支持的情况，Abascal 尤其明显（25.0% 对 10.4%）。从绝对数量看，一半的回答支持匿名帖子，四分之一表示反对（平均立场为 0.24），因此这些惩罚效应并非地板效应或天花板效应。署名使回答从支持转向反对，而中立比例仍维持在四分之一左右：署名 Abascal 时，40% 表示支持，34% 表示反对（附录 C.3，表 30）。

这种惩罚效应并不止于立场：在全部24个框架—主题—说话者单元中，点出领导人的姓名都会降低一致性、合法性和支持，并提高感知冒犯性（附录C.1，表 21）。当显示名称对应真实作者时，这种效应最小（-0.07 至 -0.09 按框架）；而对于不可信的反事实归因，这种效应最大（-0.15 和 -0.23），不过这一比较仅具有描述性，因为三个类别包含的是不同的帖子（附录 A.5，表13）。

说话者	Δ立场 [95% 置信区间]	支持度较低	Same
Sánchez	-0.100 [-0.114, -0.086]	18.6%	69.6%
Feijóo	-0.100 [-0.112, -0.087]	18.9%	69.5%
Abascal	-0.191 [-0.206, -0.175]	25.0%	64.6%

表 2：说话者归因对立场的影响。 Δ 表示在完整四元组、两种框架和所有主题中，归因与匿名条件之间差值的平均值；95%置信区间采用按消息和人格进行的双向聚类稳健估计。最后两列给出了支持度降低或保持不变的回答占比。

6.2 意识形态一致性（研究问题2）

图2a 展示了每个名称在各个引导位置上的效应，表 3则展示了一致性对比。各群体的估计值见附录 C.2，表 25。

研究问题1中的 Sánchez 和 Feijóo 整体效应几乎相同（表 2），这是不同方向反应取平均后的结果。左翼人物画像对 Sánchez 的惩罚最轻（$\Delta = -0.047$），对 Feijóo 的惩罚较重（-0.116），对 Abascal 的惩罚最重（-0.288）。右翼人物画像则颠倒了两位主要领导人的顺序：对 Feijóo 的惩罚最轻（-0.049），其次是 Abascal（-0.072），对 Sánchez 的惩罚最重（-0.129）；而中间派人格对 Sánchez 和 Feijóo 的惩罚相同（-0.123 和 -0.121），对 Abascal 的惩罚更重（-0.195）。这种变化沿量表逐渐发生（附录C.2，表24）。从位置 2 到 9，Sánchez 的效应从 0.001 变为 -0.151，Feijóo 的效应从 -0.110 变为 0.008，Abascal 的效应从 -0.342 变为 -0.037；Sánchez 的曲线在 5 到 6 之间与 Feijóo 的曲线相交，在 6 到 7 之间与 Abascal 的曲线相交，因此结果并不取决于截断点。

一致性对比按照模型在知识审计中排列领导人的顺序排列（Sánchez、Feijóo、Abascal），三者的结果均为正，且置信区间高于零（表3），其中相距最远的两位领导人 Sánchez 和 Abascal 的差异最大（0.298）。右翼人格对 Abascal 的惩罚也略高于对 Feijóo 的惩罚（0.023），这与大多数右翼人格将自身定位在 7 或 8、且更接近 Feijóo 的情况一致。在两种框架下，这一模式在全部四个

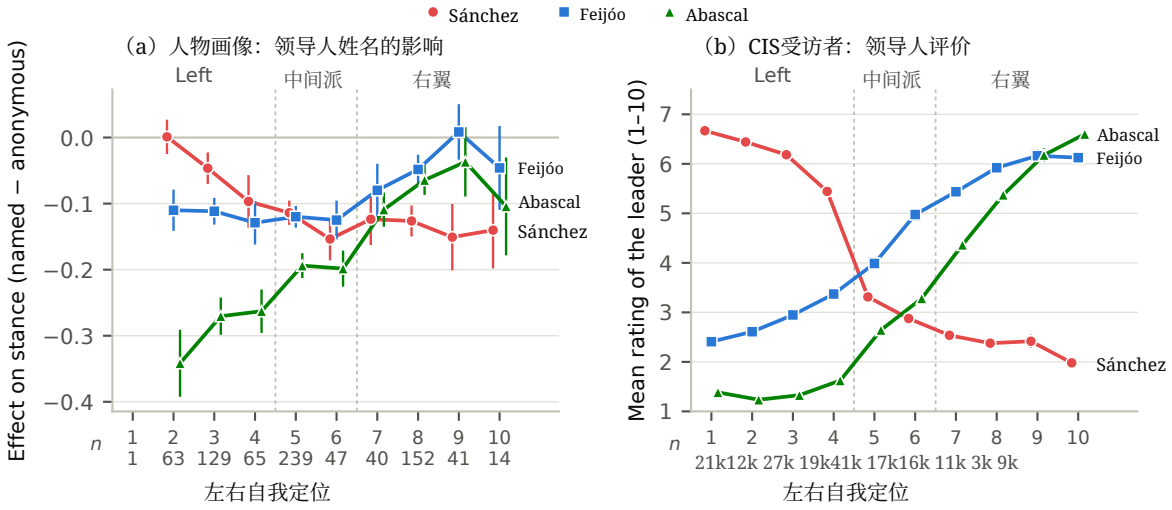

图 2：按左右自我定位划分的领导人偏好。(a) 人物画像：在每个引导定位点上、两种框架及所有主题中，说话者归因效应对立场的影响，并显示95%置信区间；定位点1（一个人物画像）未显示。(b) CIS受访者：在每个自我定位点上对每位领导人的平均评分（1–10），汇总2022年4月至2026年10月期间的48期晴雨表数据，并显示95%置信区间。n 表示每个定位点上的人物画像数量（a）和受访者数量（b）。虚线将左派（1–4）、中间派（5–6）和右翼（7–10）群体分隔开。

领导人 A – B	Left	中间派	右翼	左派 – 右翼
Sánchez – Feijóo	0.069 [0.055, 0.084]	-0.002 [-0.013, 0.009]	-0.080 [-0.097, -0.064]	0.150 [0.128, 0.172]
桑切斯—阿巴斯卡尔	0.241 [0.212, 0.271]	0.072 [0.053, 0.090]	-0.057 [-0.079, -0.036]	0.298 [0.264, 0.333]
Feijóo – Abascal	0.172 [0.152, 0.192]	0.074 [0.059, 0.089]	0.023 [0.007, 0.040]	0.148 [0.124, 0.173]

表3：立场一致性对比。前三列报告各群体内部的 $\Delta_A - \Delta_B$；正值表示该群体对A的惩罚程度低于对B的惩罚程度。最后一列为左翼群体与右翼群体之间的差异，正值表示符合一致性。单元格报告估计值 [95%置信区间]。

主题中以及全部五项结果上均成立（附录 C.2，表 27 和29）。24 组框架—主题对比中，只有两组的置信区间包含零，且均出现在 IW人物画像中（腐败主题上的 Sánchez–Feijóo 对比，移民主题上的 Feijóo–Abascal 对比）。

一致性也进一步说明了研究问题1中普遍存在的惩罚效应。除腐败议题外，将观点归因于意识形态上最接近的领导人，并未产生可检测到的惩罚效应：对于左翼人格，Sánchez的名字提高了对经济议题信息的支持（$\Delta = 0.040$ [0.026，0.055]），且未显著改变对家庭成本或移民议题信息的支持；对于右翼人格，Feijóo的名字未显著改变对任何一个主题的支持。腐败议题则是例外。在这一议题上，每个群体都会惩罚每位领导人；左翼人格对 Sánchez（−0.338）的惩罚程度与对Feijóo（−0.329）一样高。

由于引出的意识形态定位从未包含在消息评估提示中，且始终与角色的价值观相关（第 4.4 节），因此这种一致性模式与价值基础画像相关而非与明确的意识形态标签相关。

6.3 主题（RQ3）

归因惩罚的幅度因主题而异（表4，区间见附录 C.1，表23）。对所有姓名而言，腐败都是代价最高的主题：IW人物画像中的归因效应介于 −0.20 至 −0.17 之间，MFT人物画像中则介于 −0.40 至 −0.34 之间。从绝对值来看，没有作者署名时，腐败帖子接近中立（−0.04），而无论署上谁的姓名，多数人都会拒绝这些帖子（54–57%）；相比之下，家庭成本帖子在署上任何人的姓名后仍获得多数支持（附录C.3，表 30）。在经济、家庭开支和移民这几个主题上，Sánchez和Feijóo的姓名带来的效应相对较小（−0.07 至 −0.01），而Abascal的姓名在这三个主题上的惩罚均介于 −0.19 和−0.10之间，因此他的惩罚并非仅限于腐败主题。24个框架—主题—说话者估计值中，只有一个置信区间包含零：MFT人物画像中Sánchez在

	IW人物画像			MFT人物画像		
主题	Sánchez	Feijóo	Abascal	Sánchez	Feijóo	Abascal
经济与就业	-0.030	-0.055	-0.186	-0.008	-0.042	-0.182
家庭开支	-0.051	-0.057	-0.168	-0.053	-0.047	-0.132
移民与边境	-0.056	-0.030	-0.164	-0.065	-0.031	-0.101
腐败	-0.194	-0.173	-0.202	-0.395	-0.348	-0.335

表4：按主题和人物选择框架划分的说话者归因效应对立场的影响。负值表示，与同一消息的匿名版本相比，该姓名降低了支持度。置信区间见附录C.1，表23。

		CIS平均评分					晴雨表（含）
群体	受访者 Sánchez		Feijóo	Abascal	CIS 顺序	人物顺序	相同顺序
左派（1–4）	78,751	6.23	2.83	1.40	S >F > A	S > F > A	48/48
中间派 (5–6)	57,904	3.19	4.26	2.82	F >S > A	F > S > A	39/48
右翼 (7–10)	39,109	2.33	5.82	5.43	F >A > S	F > A > S	32/48

表5：CIS受访者和人物画像中，按意识形态群体划分的三位领导人排序。CIS：2022年4月至2026年10月的48期晴雨表中，各期权重相同的加权平均评分（1–10）；所有95%置信区间均在 ±0.04以内。人物画像：说话者归因效应对立场的排序（图 2a）。S：Sánchez；F：Feijóo；A：Abascal。最后一列统计排序与人物画像一致的晴雨表期数。

经济与就业主题下的估计值（$\Delta = -0.008$，95%置信区间 $[-0.024, 0.008]$）。

总体主题模式反映了研究问题2中观察到的群体特定反应（附录 C.2，表 28）。除腐败外，Sánchez和Feijóo的总体效应较小，反映出来自对立意识形态群体的惩罚，而来自其自身阵营的惩罚很少或无法检测。Abascal受到的惩罚主要来自左派和中间派人物画像；在这些主题上，他们因提及他的名字而给出的评分低 -0.14 至 -0.29，而右翼人格在家庭开支方面给他的评分接近零（-0.002）。在腐败问题上，每个群体都会惩罚所有领导人，包括其自身阵营的领导人。因此，归因效应既取决于消息说了什么，也取决于据称是谁说的。

6.4 与CIS受访者的比较（RQ4）

在整个研究期间，三个群体中的人物画像和CIS受访者对各领导人的排序都相同（表5）；在每个群体内，两位领导人之间的九项差异以及三项左派减右翼的对比，在人物画像和受访者中符号也都相同（附录D，表33）。一致性最弱的是中间派：受访者给Feijóo的评分比Sánchez高1.06分，而中间派人格几乎无法区分两人（差值为0.002，区间包含零）。

一致性并非在整个量表上都相同（图 2b 对比 2a，以及附录 D，表 34）。人物画像和受访者对领导人的排序在位置等级 1–4、7 和 8 上顺序相同。在存在差异的情况下，与处于相同位置等级的受访者相比，人物画像对 Abascal 的评价较低，或对 Sánchez 的评价较高；这种差异在 9 和 10 级最为明显：受访者对 Abascal 的评价与 Feijóo 一样高，甚至更高，而人物画像仍然更偏好 Feijóo，不过处于这两个等级的人物画像分别只有 41 个和 14 个。

一致性也取决于时间段。在右翼受访者中，从2022年4月至2024年12月的每一期晴雨表，排序都与人物画像一致；但自2025年1月以来的20期中，只有4期如此。他们对 Feijóo 与 Abascal 的评分差值（Feijóo 减 Abascal）从2022—2023年的约1个百分点，降至此后20期中有16期低于零（附录 D，图 7）。中间派受访者也呈现出同样变化，但程度较弱：截至2024年12月，每期晴雨表的排序都一致，此后20期中有11期一致。因此，与最新几期晴雨表反映的民意相比，这些人物画像更接近2022—2024年的西班牙民意。时间上的吻合与模型在2024年6月的知识截止日期相符（第 5节）。截至该月的23期晴雨表中，CIS 排序在三个群体中都与人物画像一致。这一比较无法证明知识截止日期是原因；而且，如同任何外部基准，它只能说明合成模式在排序和方向上与调查模式相似，无法说明这些人物画像是否代表西班牙公民。

7 讨论与结论

名字会改变人们对相同文本的判断，而且改变方向正如针对人类受访者的来源线索研究所预测的那样。除腐败议题外，将内容归因于意识形态上最接近的领导人，几乎不会招致惩罚，甚至毫无影响；而归因于竞争对手则会降低支持度。总体而言，Abascal 的名字带来的惩罚最大，这与语言模型对 Abascal 表达的负面情感倾向一致（Goyanes 等，2025），也与 Elbouanani 等人 (2026)记录的右翼政治人物普遍受到更严厉惩罚的现象相符。模型对 Abascal 本人发布的帖子评价也最差，尤其是在不显示姓名时。匿名阅读时，这些帖子的平均立场为 -0.35，而 Sánchez 的帖子为0.47，Feijóo 的帖子为0.28；在每个意识形态群体中，以及在两种人物选择框架下，这些帖子都得分最低（附录 C.3，表 32）。内容上的这种差异大于任何归因效应，但不会纳入 Δ，因为每条帖子都会在不同姓名标注下与自身进行比较。整体模式更多体现为排斥，而非吸引。人物画像很少因为某条消息由己方阵营署名而给予更积极的评价，但当消息由对方阵营署名时，评价就会降低；人类受访者对外党线索的反应也比对内党线索更强（Nicholson，2012）。任何姓名标注都会比匿名版本降低支持度，这表明带有归因信息的陈述可能被视为党派行为。腐败议题上的解读最为严厉：所有群体都对每位领导人予以惩罚，包括对本阵营的领导人。

这些人物画像并未被赋予明确的政治意识形态：它们是根据人物传记和价值观画像进行设定的，而其左–右位置则是在之后测定的，并且在评估消息时从未呈现。这一位置与道德基础问卷-2和英格尔哈特–韦尔泽尔维度之间存在系统性关联，其方向与人类样本中观察到的关系一致。Greco 等人(2026b)发现，基于价值观构建的人物画像对相同消息的反应各不相同。在这里，这种差异化表现出政治层面的组织性，因此，价值观画像也承载着关于谁可以说什么的立场。

与 CIS 的比较显示，这幅图景在多大程度上反映了西班牙民意，又在哪些方面有所偏离。在所有群体和每一组两两差异中，排序和方向都一致；但两者的量级完全无法比较，因为消息上的说话者归因效应与对领导人的评价是不同的数量。这种吻合也有时效限制。人物画像复现了晴雨表截至模型知识截止日期的排序，却未能反映自 2025 年以来右翼受访者转而支持 Abascal 的变化。合成受众可能带有从训练数据中继承而来的、过时的政治联想。在将人物画像用作实时竞选活动的受众之前，应像 RQ4 在此所做的那样，核查其所呈现图景的日期；其输出不应被视为当前民意。

尽管三个意识形态群体的合并领导人排序完全一致，但逐轮来看，一致性并非完美：左翼的人物顺序在48个晴雨表中有48个相符，中间派有48个中的39个相符，右翼则有48个中的32个相符。此外，仅对三位意识形态立场迥异的领导人进行排序，是一种低维度基准。这种匹配可能反映的是关于西班牙政治光谱的广泛习得关联，而非对选民行为的精细模拟，因此应将其视为有限的外部效度检验，而非调查等价性的证据。

价值导向人物画像会连贯地按照意识形态立场，对说话者是谁以及说了什么作出回应。这使它们适用于研究来源效应，但并不意味着它们可以代表选民。由此可以开展若干拓展。最直接的做法是再用两种开放权重模型重复实验，以检验这一模式是否仅适用于 gpt-oss-20b。按选举时期为帖子标注类别，可以观察说话者效应是否会在竞选活动期间增强；让人物画像回答 CIS领导人评价题项，则可以进行尺度一致的比较。对于匿名帖子仍能透露多少作者信息这一关键词下限，也可以用基于模型的合理性检验来评估每项归因。

局限性

结果来自单一模型 gpt-oss-20b。该模型的知识截止于2024年6月，而帖子则延续至2026年10月。惩罚幅度、IW人物画像集中在量表中点的现象，以及对领导人的刻画，都可能是该模型所特有的。对于之后发生的事件，模型会在不了解相关情况的前提下作出判断；而它对三位领导人的呈现，也只反映截至2024年中可获得的信息，这一点从2025年后它与右翼CIS受访者的分歧中可见一斑。目前计划再测试两种开放权重模型。

这些人物画像是生成的档案，并非真实人物；每个人物画像的意识形态都只通过一次回答来测定。我们没有测试在重新措辞问题后，这一回答是否稳定；而且，虽然量表以西班牙政治为参照，却是向大多数并非西班牙人的人物画像提出的。在IW人物画像中，41%恰好回答了5，而CIS受访者中这一比例为22%（附录B.2），因此中间派群体具有异质性，不过按位置划分的结果并不取决于分界点。

我们评估了基于 IW 和 MFT画像的价值导向人物画像，但没有将其与简单的人口统计提示或零样本基线等替代性引导方法进行基准比较。由于大语言模型中的人物画像行为高度依赖提示词，我们的结果仅刻画价值导向合成受众，并不推广至其他人物画像条件化技术。

匿名基线并非完全匿名。虽然姓名和政党标签已被移除，但帖子内容往往会暴露作者的立场。例如，即使没有姓名，Abascal 的帖子收到的评价仍远差于其他人的帖子（附录 C.3），这可能部分反映了受众对作者身份的识别。关键词标记显示，46% 的刺激材料在使用某些姓名时不可信；这一比例只是下限，而且我们没有进行基于模型的可信度评估或泄漏探测。实验设计保持完全交叉，因此反事实归因中也包含不可信的归因；而按真实、可信和不可信归因进行的划分仅具描述性，因为这三类所包含的帖子各不相同。

完全交叉设计还会产生一些不可信、在某些情况下甚至内部不一致的来源—内容组合。这些情况下尤为显著的惩罚不能单纯归因于党派反对：它们也可能反映了对来源—内容不匹配的敏感度，或维持指定角色的困难。未来研究应通过独立合理性评级和明确的矛盾度量，区分这些机制。

刺激材料是关于四个主题（经济与就业、家庭开支、移民和腐败）的原创X帖子，按关键词规则分类；人工核查显示，主主题精确率为86%，其中Abascal在家庭开支主题下、Sánchez在腐败主题下的单元格数据较少。研究设计选取了一位左翼领导人和两位右翼领导人。在此期间，PSOE左侧的政治空间没有一位固定的领导人：Yolanda Díaz于2024年6月辞去Sumar领导人一职，而各政党Frente Amplio于2026年10月12日提名Ada Colau为候选人，此时帖子已经收集完毕。该政治空间是纳入第四位领导人的理想选择，而CIS在全部48期晴雨表中都对Díaz进行了评分。提示模板和角色档案使用英语，帖子则使用西班牙语；完整的西班牙语运行留待未来研究。

CIS比较仅限于顺序和方向，因为消息上的说话者归因效应与对领导人的评级无法置于同一尺度上。CIS区间未考虑抽样设计中的聚类效应，而顺序上的一致性并不意味着这些人物画像能够代表西班牙公民。

伦理考量

本研究使用了三位公职人员的公开帖子，这些帖子依据官方 X API 的条款收集；研究还使用了 CIS 的公开微观数据，并根据其使用条件注明出处。政治人物的公开言论仍属于《通用数据保护条例》定义的个人数据。这些数据仅用于学术研究，存储于机构存储设施中，且未用于对任何私人个体进行画像。我们公开代码、配置、审计工件、帖子标识符和派生标签，但不公开帖子文本。

匿名化是一种实验控制。我们并未声称这些帖子是私密的或无法识别，反事实归因仅存在于提示词中：本文绝不会将某条帖子呈现为未曾撰写该帖的领导人的言论，附录A.2，表 9 则注明了帖子的真实作者。

主要的伦理风险在于过度解读。这些人物画像并非选民；诱导得出的群体位置是模型对价值观画像的解读；左派、中间派和右翼群体并不代表真实选民群体；与 CIS 的相似之处也仅限于领导人的排序。由于这篇预印本发表于大选前，我们强调，文中内容既不预测选票，也不评价领导人，更不衡量公众对他们的支持。归因惩罚描述的是单个模型的行为，并不能说明公众的看法。

致谢

本研究部分由以下项目资助：ELENI：智能叙事地方战略，由CDTI资助（IDI-20260067）；多模态与多传感器数据中心型人工智能模型（ANNOTATE-MULTI2），项目编号 PID2024-156022OB-C32，由

MICIU/AEI/10.13039/501100011033资助，并由欧洲区域发展基金/欧盟资助。

作者使用了 AI 助手（Claude, Anthropic）进行语言编辑、撰写和检查分析及图表脚本，并非正式地审阅了部分角色标记匹配样本（附录 A.5）。研究设计、分析和论述均由作者完成，所有文本也都经过作者审阅。

参考文献

Alessia Antelmi、Alessia Galdeman、Lucio La Cava、Arianna Pera 和 Giovanni Da San Martino。2026年。大型语言模型中的政治说服与支持。预印本，arXiv:2606.05961。

Lisa P. Argyle、Ethan C. Busby、Nancy Fulda、Joshua R. Gubler、Christopher Rytting 和 David Wingate。2023年。由一而众：使用语言模型模拟人类样本。《政治分析》，31(3):337–351。

Mohammad Atari、Jonathan Haidt、Jesse Graham、Sena Koleva、Sean T. Stevens 和 Morteza Dehghani。2023年。超越 WEIRD 的道德观：道德的定律网络如何因文化而异。《人格与社会心理学杂志》，125(5):1157–1188。

Cristina Aybar、Virgilio Pérez 和 Jose M. Pavía。2024年。量表很重要：揭示李克特量表对政治自我定位的影响。《质量与数量》，58(4)：3725–3746。

Michael Barber 和 Jeremy C. Pope。2019年。政党是否凌驾于意识形态之上？厘清美国的政党与意识形态。《美国政治学评论》，113(1)：38–54。

Jan Batzner、Volker Stocker、Bingjun Tang、Anusha Natarajan、Qinhao Chen、Stefan Schmid 和 Gjergji Kasneci。2025。人物角色归谁所有？大语言模型研究中的合成人物角色实验及其透明化路径。载于*AAAI/ACM*人工智能、伦理与社会大会论文集，第8卷，第343–354页。

Pietro Bernardelle、Leon Fröhling、Stefano Civelli、Ric- cardo Lunardi、Kevin Roitero 和 GianlucaDemar- tini。2025。利用合成人物角色绘制并影响大语言模型的政治意识形态。载于*2025*年*ACM*万维网大会伴随论文集，第864–867页。ACM。

James Bisbee、Joshua D. Clinton、Cassy Dorff、Brenton Kenkel 和 Jennifer M. Larson。2024年。人类调查数据的合成替代品？大语言模型的风险。*PoliticalAnalysis*，32(4)：401–416。

Julien Boelaert、Samuel Coavoux、Étienne Ollion、Ivaylo Petev 和 Patrick Präg。2025。机器偏见。生成式语言模型如何回答民意调查？《社会学方法与研究》，54(3)：1156–1196。

John G. Bullock。2011。知情选民中的精英对公众舆论的影响。《美国政治学评论》，105(3)：496–515。A. Colin Cameron、Jonah B. Gelbach 和 Douglas L. Miller。2011。多路聚类下的稳健推断。《商业与经济统计杂志》，29(2)：238–249。

社会学研究中心。2026a。2026年10月晴雨表（研究3582）。访问日期：2026年11月9日。

社会学研究中心。2026b。2022年4月至2026年10月月度晴雨表（研究3359至3582）。微观数据文件，访问日期：2026年11月9日。

Geoffrey L. Cohen。2003。党派高于政策：群体影响对政治信念的主导作用。《人格与社会心理学杂志》，85(5)：808–822。

Saloni Dash、Amélie Reymond、Emma Spiro 和 Aylin Caliskan。2026。赋予角色的大语言模型表现出类似人类的动机性推理。载于计算语言学协会会议论文集：*ACL 2026*，第12043–12069页，圣迭戈，加利福尼亚州，美国。计算语言学协会。

JamesN.Druckman、ErikPeterson 和 Rune Slothuus。2013 年。精英党派极化如何影响公众意见的形成。《美国政治学评论》，107(1)：57–79。

Akram Elbouanani、Evan Dufraisse 和 Adrian Popescu。2025。通过面向目标的情感分类分析大语言模型中的政治偏见。载于计算语言学协会年会论文集：*ACL 2025*，第 15476–15505 页，奥地利维也纳。计算语言学协会。

Akram Elbouanani、Aboubacar Tuo 和 Adrian Popescu。2026年。用于审计大语言模型偏见的可扩展基于实体的框架。载于计算语言学协会年会论文集：*ACL 2026*，第 33823–33852 页，美国加利福尼亚州圣迭戈。计算语言学协会。

Shangbin Feng、ChanYoung Park、Yuhan Liu 和 Yulia Tsvetkov。2023年。从预训练数据到语言模型再到下游任务：追踪导致不公平自然语言处理模型的政治偏见轨迹。载于计算语言学协会第*61*届年会论文集（第*1*卷：长篇论文），第11737–11762页，加拿大多伦多。计算语言学协会。

E. C. Fieller、H. O. Hartley 和 E. S. Pearson。1957年。秩相关系数检验。一。*Biometrika*，44(3–4)：470–481。

Federico Germani 和 Giovanni Spitale。2025。来源框架会在大语言模型中引发系统性偏差。*ScienceAdvances*，11(45):eadz2924。

Manuel Goyanes、Adrián Domínguez-Díaz 和 Luis de Marcos。2025。14种人工智能大语言模型的人格特质、政治倾向、意识形态及其对政治领导人的情感倾向。*Human Behavior and Emerging Technologies*，2025:5761832。

Jesse Graham、Jonathan Haidt、Sena Koleva、Matt Motyl、Ravi Iyer、Sean P. Wojcik 和 Peter H. Ditto。2013。道德基础理论：道德多元主义的实用有效性。载于 Patricia Devine 和 Ashby Plant 编，实验社会心理学进展，第47卷，第55–130页。Academic Press。

Jesse Graham、Jonathan Haidt 和 Brian A. Nosek。2009。自由派和保守派依赖不同的道德基础体系。*JournalofPersonality andSocial Psychology*，96(5):1029–1046。

Candida M. Greco、Lucio La Cava 和 Andrea Tagarelli。2026a。植根于文化的角色设定在大语言模型中的应用：社会心理价值框架下的特征描述与一致性。预印本，arXiv:2601.22396。

Candida Maria Greco、Víctor Yeste、Andrea Tagarelli 和 Paolo Rosso。2026b。“我自己的道德观，我自己的思维”：特朗普消息中角色设定条件下的道德与文化接受。载于 计算语言学协会论文集：*EMNLP 2026* 研究成果，第 51567–51595 页，匈牙利布达佩斯。计算语言学协会。Ronald Inglehart 和 Christian Welzel。2005 年。现代化、文化变迁与民主：人类发展序列。剑桥大学出版社，英国剑桥。J. Matias Kivikangas、Belén Fernández-Castilla、Simo Järvelä、Niklas Ravaja 和 Jan-Erik Lönnqvist。2021 年。道德基础与政治取向：系统性综述与荟萃分析。《心理学通报》，147(1):55–94。Andy Liu、Mona Diab 和 Daniel Fried。2024 年。评估角色设定引导生成中的大型语言模型偏见。载于计算语言学协会论文集：*ACL 2024* 研究成果，第 9832–9850 页，泰国曼谷。计算语言学协会。Javier Lorenzo-Rodríguez 和 Mariano Torcal。2022 年。Twitter 与情感极化：关注西班牙的政治领导人。*SouthEuropeanSociety andPolitics*，27(1):97–123。

Marlene Lutz、Indira Sen、Georg Ahnert、Elisa Rogers 和 Markus Strohmaier。2025。提示词塑造人格：针对大型语言模型社会人口统计学人格提示的系统性评估。载于 计算语言学协会论文集：*EMNLP 2025* 研究成果，第 23212–23237 页，中国苏州。计算语言学协会。

Fabio Motoki、Valdemar Pinho Neto 和 Victor Rodrigues。2024。比人类更像人类：衡量 ChatGPT 的政治偏见。《公共选择》，198(1–2)：3–23。

Stephen P. Nicholson。2012。极化线索。《美国政治学杂志》，56(1)：52–66。

OpenAI。2025。gpt-oss-120b 和 gpt-oss-20b 模型卡。预印本，arXiv:2508.10925。

Seong-Jin Park 和 Kang-Min Kim。2025。衡量并缓解大语言模型中的媒体机构名称偏见。载于 *2025* 年自然语言处理实证方法会议论文集，第 29778–29797 页，中国苏州。计算语言学协会。

Andres Reiljan。2020。欧洲政党体系中的「党派间恐惧与厌恶」：情感极化。《欧洲政治研究杂志》，59(2):376–396。

David Rozado。2024年。大语言模型的政治偏好。*PLOS ONE*，19(7):e0306621。

Shibani Santurkar、Esin Durmus、Faisal Ladhak、Cinoo Lee、Percy Liang 和 Tatsunori Hashimoto。2023年。语言模型反映的是谁的观点？载于第*40*届国际机器学习会议论文集，第202卷，《机器学习研究论文集》，第29971–30004页。PMLR。Joseph Suh、Erfan Jahanparast、Suhong Moon、Minwoo Kang 和 Serina Chang。2025。基于规模化调查数据微调语言模型，以预测公众意见的分布。载于计算语言学协会第*63*届年会论文集（第*1*卷：长篇论文），第21147–21170页，奥地利维也纳。计算语言学协会。Mariano Torcal 和 Emily Carty。2026年。西班牙的意识形态极化及党派争端的作用（1987–2024）。*Panorama Social*，43：75–88。Mariano Torcal 和 Josep M. Comellas。2022。政治不稳定与冲突时期的情感极化：比较视角下的西班牙。*SouthEuropean Society and Politics*，27(1)：1–26。Sebastián Vallejo Vera 和 Hunter Driggers。2025年。大语言模型作为标注员：政党线索对大语言模型标注决策的影响。人文与社会科学传播，12：1530。

Leah von der Heyde、Anna-Carolina Haensch 和 Alexander Wenz。2026 年。民众之声，人工智能之声？使用大语言模型估计德国投票选择。社会科学计算评论，44(3)：549–571。

Angelina Wang、Jamie Morgenstern 和 John P.Dickerson。2025 年。取代人类参与者的大语言模型可能会有害地歪曲并扁平化身份群体。《自然·机器智能》，7(3)：400–411。

Albert Webson 和 Ellie Pavlick。2022 年。基于提示词的模型真的理解其提示词的含义吗？载于计算语言学协会北美分会 *2022* 年会议论文集：人类语言技术，第 2300–2344 页，美国西雅图。计算语言学协会。

Franziska Weeber、Vera Neplenbroek、Jan Batzner 和 Sebastian Padó。2026 年。一种人格，多种线索，不同的结果：社会人口统计线索如何影响大语言模型个性化。载于计算语言学协会第 *64* 届年会论文集（第 *1* 卷：长篇论文），第 44892–44921 页，美国加利福尼亚州圣迭戈。计算语言学协会。

A 刺激材料、提示与归因检查

A.1 语料库和筛选条件

表6 按领导人列出了质量和匿名化检查。首先清理文本（解码 HTML 实体、移除网址、合并空白字符），阈值沿用前一篇论文的标准：少于 40 个字符或少于 7 个词。无需列出其中完全符合条件的低信息量文本，因为这类文本本就少于 7 个词；截断标记也不适用，因为 X API 会返回完整文本。匿名化后，接下来进行两项检查：身份泄露（例如因拼写错误或缺少空格，领导人或政党的名称仍然留在文本中）和质量退化（少于 40 个字符、去除占位符后少于 7 个词，或在 80 个字符以内出现超过 3 个占位符）。如果某条帖子的匿名化文本（忽略大小写）与较早发布且通过所有检查的帖子重复，则判定为重复帖文。一条帖子可能未通过多项检查；大多数质量退化的帖子原本就太短，只有 24 条仅因质量退化而未通过（例如活动公告，如“4 月 28 日星期五在 #巴达霍斯 [竞选活动_标签]”）。被排除的帖子主要是问候语、口号和纯图片帖文，326 条中只有 8 条涉及四个主题之一。

	Sánchez	Feijóo	Abascal	All
原始帖文	3,295	3,032	1,298	7,625
少于40个字符	31	90	90	211
少于7个词	31	90	137	258
身份泄露	0	0	2	2
文本降级	36	109	154	299
重复文本	7	6	10	23
通过所有检查	3,252	2,915	1,132	7,299

表 6：原始帖子质量与匿名化检查。一条帖子可能未通过多项检查。

表7列出了主题规则及其结果。显著性指在所选时间段内有可用表格的 46 期晴雨表中，将该议题列为西班牙三大问题之一的受访者比例的平均值。（8 月，以及大选前后的 2023 年 6 月和 9 月均未发布晴雨表；2023 年 5 月和 2024 年 12 月的表格则无法解析。）在排名之前，我们排除了将政治本身列为问题的类别（政府与政党、政治人物行为、缺乏共识），因为这些类别并未指向任何政策领域，且与本实验所测量的对政治人物的评价有所重叠。宽泛类别（价值观危机、对未来的不确定性）和非实质性答案也予以排除。下载数据之前，已将经济、失业和工作质量合并为一个主题，住房、能源和价格也同样合并；逐一撤销这些合并后，排名仍然相同，唯一的例外是将经济拆分为三个类别会使腐败升至第七位。排名第六的议题是教育，其平均显著性为 7.1%。

帖子依据西班牙语关键词库被归入五个议题，匹配前先去除网址和重音符号，并预先删除少数短语，因为它们包含具有其他含义的主题词（“卫生封锁线”、“蒙克洛亚宫的房客”、火灾和洪水中撤离的住宅）。在对80条匹配帖子的抽查中（每个保留主题20条，一名标注者），经济、家庭开支、移民和腐败四个主题各20条帖子中，匹配词以该主题含义使用的分别有19、18、20和20条，帖子主要内容关于该主题的分别有17、16、19和17条。三处词义错误已在关键词库中修正；其余错误是关键词无法区分的顺带提及（法律清单中的“empleo”或“vivienda”，以及关于洪水或加沙战争的帖子中的“corrupción”）。

主题	CIS %	Sánchez	Feijóo	Abascal	保持不变
经济	63.2	305	100	30	yes
家庭	22.9	123	95	11	yes
健康	14.6	92	56	4	no
移民	13.8	50	40	57	yes
腐败	8.3	27	128	117	yes
多个主题		55	33	32	
无主题		2,643	2,580	1,047	

表7：主题规则：CIS民意调查中最突出的五个议题（CIS %：2022年4月至2026年10月间提及该议题的受访者平均比例）及与每个议题匹配的原创帖子，均为质量检查前的数据。只有当每位领导人至少有十条帖子时，才保留该议题；Abascal 的健康议题未达标。匹配两个议题或一个议题都未匹配的帖子，均未用作刺激材料。

A.2 匿名化

表 8 列出了每个占位符所替换的内容。其他政治人物的姓名予以保留（Ábalos、Ayuso、Puigdemont、特朗普）：替换它们会使许多帖子变空，尤其是关于腐败的帖子，而附录 A.5 中的角色标记检查涵盖了它们暴露作者身份的情形。单独的名字也予以保留，因为“Pedro”大多指别人；关于内容、地点和日期的话题标签（#DANA、#COP28、#23J）以及政党形容词（“socialistas”）同样保留。通过每一项检查的帖子中约有一半包含占位符（Sánchez 为 3,252 条中的 1,348 条，Feijóo 为 2,915 条中的 1,516 条，Abascal 为 1,132 条中的 747 条）；[政治_行动者] 出现在其中的 157、661 和 283 条中，而在 Feijóo 和 Abascal 的帖子中，它几乎总是代表 Sánchez。表 9 展示了模型所看到的一条帖子。

A.3 提示模板

图3 展示了一次调用中的系统提示和用户提示。每次调用都会向一个人格呈现同一主题的十条匿名化帖子，并设定一种条件；四种条件的差异仅在于条件行。角色档案由该人格的传记及其框架中的价值观构成，格式沿用前一篇论文。附录B.2 中的意识形态调用使用相同的系统提示和角色档案，随后附上自我定位问题和一个单字段模式。

A.4 知识审计

在实验开始前，模型分别针对三位领导人以及三个对照名称中的每一个进行了20项请求：两个虚构但可信的西班牙姓名和一个不透明代码（表10；共120项请求，使用与

占位符	替换
[政治_行动者]	三位领导人：全名，姓氏“佩雷斯-卡斯特洪”，“桑切斯主义”、“桑切斯主义者”及其自己的账号
[政党]	PSOE、人民党、呼声党及其地区分支机构（PSC、PSdeG、PSPV……），“人民党”、“社会党”、PPE；政党账号及欧洲政党家族
[竞选_标签]	包含领导人或政党名称的话题标签名称（#VotaPSOE, #InvestiduraFeijóo, #SánchezDimisión）和 54 个政党从……中人工挑选的口号使用至少三次的 203 个话题标签次（#EsElMomento, #ElGobiernoDeLaGente，#SentidoComún）
[用户]	除27以外的所有其他账号使用普通名称的公共机构名称（@Congreso_Es 变为“众议院”，@NATO 变为“北约”）

表8：匿名化占位符。

原文（Feijóo，2023年1月31日；末尾的图片链接已在清理步骤中移除） 末尾的图片链接已在清理步骤中移除） 佩德罗·桑切斯夸耀就业岗位的创造以及 经济增长。唯一的事实是，西班牙 处于欧盟末位。
匿名化（在四种条件下均显示） [政治_行动者]夸耀创造就业机会并 实现经济增长。唯一的事实是 西班牙在欧盟中排名垫底。
条件行，无作者 在此条件下未提供说话者身份。请评估 每条消息都不预设其作者是谁。
条件行，归因于（佩德罗·桑切斯、阿尔贝托 努涅斯·Feijóo 或圣地亚哥·阿巴斯卡尔） 在此实验条件下，以下每条消息都 归因于佩德罗·桑切斯。评估消息时，请将此归因视为既定事实 评估消息时。

表9：一个刺激材料和四个条件行。角色标记规则将这条帖子标记为不太可能以 Sánchez 的名义发布，因为它是在攻击他。

系统
请扮演所提供角色档案中描述的人物。以此人的方式评估西班牙语政治信息，并将显示的发言者归属视为实验背景的一部分。将显示的归属视为既定事实，不要质疑其在历史上是否准确。若未注明说话者，不要根据风格、主题或外部知识推断说话者。不要试图查明或提及隐藏的真实作者。将实质立场与修辞语气区分开来。仅返回有效的 JSON，不要使用 Markdown。

用户[角色档案]调查受访者概况。传记： <传记>IW人物画像：英格尔哈特—韦尔策尔空间中的文化定位：生存—自我表达 = <x>，传统—世俗 = <y>。解读提示：生存—自我表达得分越低，越接近生存／物质安全取向；得分越高，越接近自我表达取向。传统—世俗得分越低，越接近传统价值观；得分越高，越接近世俗理性价值观。MFT人物画像： [道德档案] 根据道德基础问卷-2得出的道德档案，采用 1..5 量表。数值越高，表示该陈述越能描述此人。道德基础问卷-2基础维度得分： <六项基础维度得分>。下方提供了道德基础问卷第二版题目直接作答。请将其作为此人道德直觉的具体依据。道德基础问卷第二版作答量表：完全不能描述我（1）；略微描述我（2）；中等程度描述我（3）；相当符合地描述我（4）；非常符合地描述我（5）。道德基础问卷-2题目回答：1. <题目 1> 回答： <1-5> ... 36. <题目 36> 回答： <1-5>

[政治背景]
Th这些消息涉及西班牙当前的政治议题。无作者：在此条件下不提供说话者身份。评估每条消息时，不要假设作者是谁。归因：在此实验条件下，以下每条消息均归因于<佩德罗·桑切斯 |阿尔贝托·努涅斯·Feijóo | 圣地亚哥·阿巴斯卡尔>。评估消息时，请将此归因视为既定事实。

[评估规则]
- 立场必须为支持、中立或反对，并应主要反映实质性立场。- 一致性： 1 =强烈反对， 5 = 强烈赞同。- 合法性：从此人格的角度看， 1 = 完全不合法， 5 = 完全合法。- 冒犯性： 1 =不具冒犯性， 5 = 极具冒犯性。- 支持： 1 =绝对不会分享， 5 = 绝对会按原文分享该消息。- 立场和一致性须保持一致：支持通常对应 3..5，反对对应 1..2，中立对应 2..3。- 情绪_反应最多只能包含 20 个词。- 简短_理由最多只能包含 60 个词。- 独立评估每条消息，并且每个推文_id 恰好返回一条回复。

[消息]
1. [推文_编号=<编号>]消息： <
匿名化帖子>.....十条消息.....

[输出 JSON 模式]{"responses": [{"tweet_编号": "<编号>", "立场": "支持|中立|反对", "一致性": 1, "合法性": 1, "冒犯性": 1, "支持": 1, "情绪_反应": "简短文本", "简要_理由": "简短文本"}, ...]}

图 3：提示模板。尖括号标示每次调用中填入的部分；缩进的备选项展示了两种角色格式和两种条件行。解码：温度 0，随机种子42，4,096 个输出词元，JSON 模式。

实验）。对于每个姓名，20项请求由四个相互独立的探测系列中的每个系列各包含五个语义等价的提示变体组成：识别、政党与政治角色、左右意识形态以及议题立场。意识形态探测使用 1–10 量表，其中 1 表示政治左翼，10 表示政治右翼。议题立场探测进一步使用四个 1–10 量表：国家—市场取向（1 =更多政府干预； 10 = 更强的市场取向和减税）、住房政策（1 = 更多监管和租户保护；10 =更强调住房供给、市场机制和住房所有权）、移民限制程度（1 = 更开放或更具人道主义立场； 10 = 更严格的立场，强调边境管控），以及制度信任（1 =反建制论述或低制度信任； 10 =制度延续性或高制度信任）。模型还在其中若干维度上区分了三位领导人；例如，他们在移民限制程度上的中位数得分分别为 3、6 和 10。意识形态和政策立场得分描述的是模型的内部表征，并不被视为关于这些政治人物的真实情况。审计检验的是这些姓名是否会激活稳定且彼此 distinct 的政治身份；它并不估计发言者归属的影响。

Name	受认可度	政党	左–右
佩德罗·桑切斯	5/5	西班牙工人社会党 5/5	3 (3–3)
阿尔贝托·努涅斯·费霍	5/5	人民党 5/5	8 (7–8)
圣地亚哥·阿巴斯卡尔	5/5	呼声党 5/5	9 (9–10)
达里奥·文托萨尔·布鲁梅特	0/5	none	–
露西亚·蒙塔尔沃·森德拉尔	0/5	none	–
不透明代码	0/5	none	–

表 10：基于每个姓名 20项请求的知识审计：四类探测各包含五种提示词变体。识别和政党认同根据五项相关探测报告，左—右位置则报告五项意识形态探测结果的中位数（范围）。最后三个姓名是对照项：两个虚构姓名，以及代码 ES_POLITICIAN_CONTROL_QZ17。

A.5 角色标记与不合理归因

匿名化会删除姓名，但不会消除立场。表 11 列出了五组角色标记，这些标记表明帖子不太可能由另一个名字所代表的人发布。匹配方式与主题关键词相同（删除网址和重音符号，并转换为小写），且仅在各组所适用的作者所发帖子中进行；地区及外国政府和外国总理在匹配前会被剔除。表12按主题和真实作者列出了被标记的刺激材料。对随机抽取的样本进行非正式阅读（并非正式标注）后发现，准确率较高：在最后修订前阅读的 60 条标记帖子中，有 54 条体现了作者的立场；修正了 6 处错误后，之后阅读的 16 条标记主题帖子全部如此。召回率较低：Sánchez 的大多数未标记主题帖子仍以政府的口吻发言，却没有使用标记词（“我们将拨款 22.75 亿欧元”）；Feijóo 和 Abascal 的此类帖子中，约有一半仍表达反对。因此，这些标记仅构成一个下限。若据此筛选，家庭开支主题将只剩下 7 条 Abascal 的帖子，腐败主题则会留下 Feijóo 和 Abascal 的帖子各 26 条和 23 条；而且仍会保留许多不可信的帖子，因此研究设计仍采用完全交叉设计。表 13报告了显示名称与真实作者相对应时的归因效应；由于三个类别包含的帖子各不相同，这一比较仅为描述性比较。

A.6 完整性

表 14按框架和主题列出了在四种条件下观察到的人物画像–帖子配对。每个单元格的完成率均超过 99.95%；缺失的四元组来自重试后 JSON 验证仍未通过的调用。

B 人设与意识形态

B.1 人格选择

人格池取自Greco 等人 (2026a)的资源。它包含根据世界价值观调查的十个变量生成的概况——宗教信仰、育儿价值观、道德可接受性、信任与社会资本、政治参与、民族自豪感、幸福感、性别平等、物质主义取向以及对外群体的态度——并附有传记、英格尔哈特–韦尔策尔坐标和道德基础问卷-2 得分。对于每个主题和理论框架，我们都在相应的价值空间中定义了高、中、低目标概况（表15）。我们采用了Greco 等人 (2026b)在研究特朗普政治消息的受众反应时使用的相同人物画像排序和筛选程序，只调整了政治议题与相关 IW 维度或道德基础之间的对应关系。对于每种主题–概况组合，我们保留排名最高的 50 个人物画像，且同一主题的不同层级之间不重复使用人物画像。高、中、低描述的是在

标记组	发帖者	不太可能出现于	示例标记
政府声音	Sánchez	Feijóo、Abascal	“政府”、部长、部长会议、欧盟主席国任期、峰会、“对应方”
反政府	Feijóo、Abascal	Sánchez	“政府”、部长、“Sánchez”、“桑切斯主义”、辞职、不信任动议
反对右翼	Sánchez	Feijóo、Abascal	“极右翼”、“右翼”、“反动派”
反对左翼	Feijóo、Abascal	Sánchez	“左翼”、“共产党人”、“进步派”
反对人民党	Abascal	Feijóo	“人民党与工人社会党”、“两党制”、“右翼小党”、“人民党人士和社会主义者”

表11：角色标记组。标记仅与适用作者的帖子相匹配，并将该帖子标记为不太可能出现于第三列所列名称之下。

主题	Sánchez	Feijóo	Abascal	All
经济	83/304	33/99	10/30	126/433
家庭	41/123	60/94	4/11	105/228
移民	17/50	22/40	19/56	58/146
腐败	15/27	101/127	91/114	207/268
All	156/504	216/360	124/211	496/1,075

表12：按主题和真实作者分类，至少带有一个角色标记的刺激材料（带标记数/总数）：Sánchez 的帖子占 31%，Feijóo 的占 60%，Abascal 的占 59%，总体占 46%。

框架	归因	对	Δ 立场	95%置信区间
IW	真实作者	161,230	-0.072	[-0.084, -0.061]
IW	可信	224,683	-0.131	[-0.144, -0.119]
IW	不可信	97,777	-0.145	[-0.161, -0.129]
MFT	真实作者	161,240	-0.089	[-0.106, -0.072]
MFT	可信	224,683	-0.149	[-0.167, -0.132]
MFT	不可信	97,797	-0.233	[-0.260, -0.205]

表13：显示名称与真实作者是否对应对立场的说话者归因效应。当角色标记表明帖子以该名称发布时，反事实归因不可信。

	IW				MFT			
	经济	家庭	移民	腐败	经济	家庭	移民	腐败
对	64,950	34,200	21,900	40,200	64,950	34,200	21,900	40,200
完成	64,950	34,200	21,890	40,190	64,940	34,200	21,900	40,200
%	100.00	100.00	99.95	99.95	99.99	100.00	100.00	100.00

表14：按框架和主题划分的完整四条件四元组。

相应理论值空间中的位置，且绝不会在提示词中展示给模型。

图4展示了相对于完整人格池的最终 IW选择。表16通过 50 个人物画像中各道德基础得分的均值和标准差，总结了相应的MFT选择。

B.2 引导式自我定位

791个不同的人物画像（441个IW、350个MFT）均作答一次。反应提示词中的系统提示和角色档案之后，是以下问题：一般来说，在讨论西班牙政治时，通常会使用“左派”和“右翼”这两个说法。在*1*到*10*的量表上，*1*表示“最左倾”，*10*表示“最右倾”，就西班牙政治而言，您会把自己定位在哪个位置？ 这是对CIS晴雨表题项的英文改编。该量表以西班牙政治为参照，因为大多数人物画像并非西班牙人，而且与反应提示词不同，这些人物画像并未被设定为西班牙居民。政党和领导人并未被提及，因为领导人偏好是RQ4中用于比较的结果。设置与反应提示词中的设置相同，791份答案均为1到10之间的有效整数（表17）。许多IW人物画像将自己定位在5，即1至10量表中两个中间选项里较低的那个：441个IW人物画像中有180个（41%）恰好回答5；相比之下，350个MFT人物画像中有59个（17%）如此作答，而2026年10月晴雨表中的受访者这一比例为22%。

B.3 价值观与自我定位

表19 列出了各画像层级上的意识形态群体，表18则展示了提示词中的自我定位与各价值维度之间的相关性。若某一画像具有明确的政治倾向，则其与引导出的意识形态群体高度相关：在两种框架中，包容型和限制型移民画像大多分别呈左倾与右倾，而 MFT经济与家庭开支画像则对比了关怀与

主题	层级	IW画像	MFT画像
经济、就业	high	物质保障较弱理性政策成分	功绩与经济贡献
	mid	没有鲜明主题的中心档案方向	没有明显两极分化的经济取向
	low	自我表达与工作质量取向	再分配与社会保护
家庭开支	high	可负担性与基本物质保障	可负担性与基本需求保障
	mid	没有鲜明主题的中心档案方向	未明显两极分化的家庭成本取向
	low	居住选择、自主性和自我表达	市场配置和个人责任
移民、边境	high	包容、开放和多元主义取向	人道主义包容
	mid	没有明确主题的中心档案方向	没有高度两极化的移民议题取向
	low	限制、安全与传统秩序	限制与边境管控
腐败	high	程序与制度问责	强制度问责敏感度
	mid	没有强烈主题的中心档案方向	中度制度问责敏感度
	low	传统或个人化权威取向	制度问责敏感度较弱敏感度（并非支持腐败）

表15：T 每个框架中用于为每个单元格选取50个人物画像的主题条件化画像层级。这些层级将人物画像 er置于 框架的价值空间内；它们不是意识形态标签。

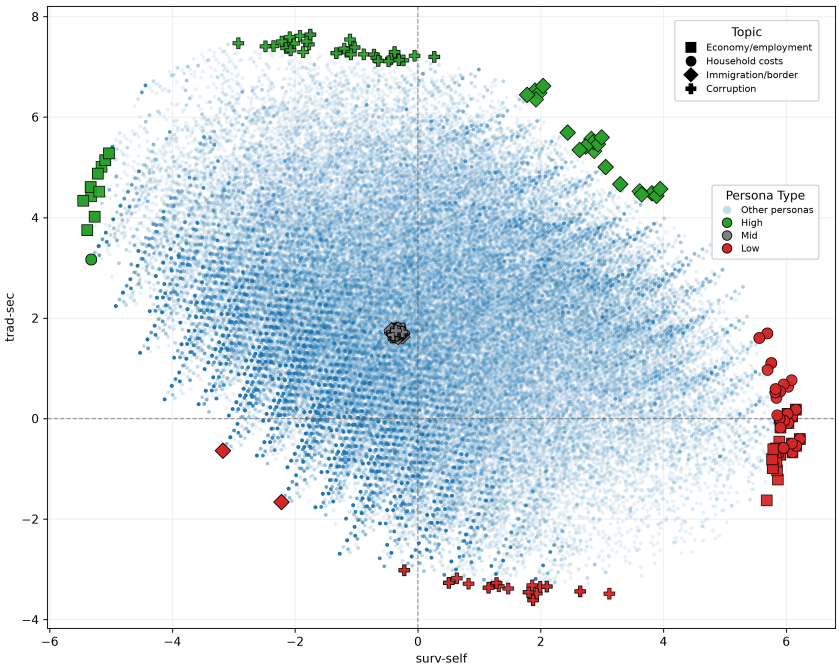

图4：为每个主题–画像组合选取的 Inglehart–Welzel 人物画像。背景点表示 IW空间中的完整人格池，突出显示的标记表示为每个主题和画像选取的50个人物画像。颜色表示主题条件化画像（高、中 和 低），标记形状表示主题。

概况	Care	平等	比例原则	忠诚	权威	纯洁
经济/就业						
High	1.17 ± 0.21	1.20 ± 0.13	4.79 ± 0.13	2.71 ± 0.61	3.47 ± 0.47	2.43 ± 0.50
Mid	3.26 ± 0.43	2.73 ± 0.35	3.45 ± 0.35	2.87 ± 0.39	3.11 ± 0.37	2.80 ± 0.26
Low	5.00 ± 0.02	4.70 ± 0.32	2.32 ± 0.38	2.93 ± 0.43	2.99 ± 0.55	2.92 ± 0.64
家庭开支						
High	5.00 ± 0.02	4.70 ± 0.32	2.32 ± 0.38	2.93 ± 0.43	2.99 ± 0.55	2.92 ± 0.64
Mid	3.26 ± 0.43	2.73 ± 0.35	3.45 ± 0.35	2.87 ± 0.39	3.11 ± 0.37	2.80 ± 0.26
Low	1.17 ± 0.21	1.20 ± 0.13	4.79 ± 0.13	2.71 ± 0.61	3.47 ± 0.47	2.43 ± 0.50
移民/边境						
High	4.86 ± 0.29	3.01 ± 0.48	3.49 ± 0.45	1.93 ± 0.21	1.80 ± 0.27	2.27 ± 0.42
Mid	3.26 ± 0.43	2.73 ± 0.35	3.45 ± 0.35	2.87 ± 0.39	3.11 ± 0.37	2.80 ± 0.26
Low	1.77 ± 0.47	1.91 ± 0.56	4.25 ± 0.41	4.67 ± 0.24	4.79 ± 0.28	3.17 ± 0.51
腐败						
High	4.14 ± 0.98	4.09 ± 0.45	4.57 ± 0.39	3.42 ± 0.64	4.53 ± 0.38	3.57 ± 0.64
Mid	3.26 ± 0.43	2.73 ± 0.35	3.45 ± 0.35	2.87 ± 0.39	3.11 ± 0.37	2.80 ± 0.26
Low	3.40 ± 0.80	1.92 ± 0.44	3.16 ± 0.32	2.14 ± 0.49	2.06 ± 0.36	2.08 ± 0.40

表16：针对每个主题条件化画像选出的50个人物画像的道德基础问卷-2基础维度得分（1–5）。数值以均值 ± 标准差的形式报告。单元格底色与均值得分成正比：浅灰色表示较低值，深灰色表示较高值。

	左中		右翼	均值
IW人物画像（441）	27.2%	44.4%	28.3%	5.40
MFT人物画像（350）	39.4%	25.7%	34.9%	5.36
CIS，2026年10月（5,024）	45.4%	29.2%	22.0%	4.51

表17：引导得出的人物画像与CIS受访者的左右自我定位（社会学研究中心，2026a）。CIS受访者比例以全体受访者为基数；3.4%回答“不知道”或未作答，而人物画像没有这一选项。

价值维度	ρ	95%置信区间
*IW*人物画像		
生存／自我表达	−0.60	$[-0.66, -0.54]$
传统／世俗理性	−0.06	$[-0.16, 0.03]$
*MFT*人物画像		
Care	−0.87	$[-0.89, -0.84]$
平等	−0.67	$[-0.73, -0.61]$
比例原则	0.67	$[0.61, 0.73]$
忠诚	0.43	$[0.34, 0.51]$
权威	0.59	$[0.52, 0.66]$
纯洁	0.21	$[0.10, 0.31]$

表18：提示词中呈现的各价值维度与引出的自我定位（1–10）之间的斯皮尔曼相关，每个人物画像作为一个观测值（441个IW，350个MFT）。负值表示分数越高，定位越偏左。置信区间根据费舍尔的 z 方差为 $1.06/(n-3)$(Fieller 等人，1957年）。

平等与比例原则完全落在同一侧。无论在哪个框架中，腐败特征都未按意识形态呈现出清晰的单调区分，而中等水平大多处于中间位置。在MFT人物画像中，定位遵循基础得分：关怀和平等得分越高，定位越偏左；比例原则、权威和忠诚得分越高，定位越偏右；而纯洁的关联较弱。这些方向与Atari 等人(2023年在19个国家的样本中发现的结果一致；这些样本回答了类似的10点评分左右项目：右翼定位与关怀和平等负相关，与比例原则、忠诚、权威和纯洁正相关，其中纯洁的关联最弱。这些估计值来自一个纳入全部六项基础

将这些方向放在一起比较，因此只有方向具有可比性。在 IW人物画像中，其位置遵循生存—自我表达坐标，却几乎不遵循传统—世俗理性坐标。由于某些 MFT 单元格包含相同的人物画像（附录 B.1），它们完全偏向一方的划分主要表明，模型将高关怀和平等解读为左派，将高比例原则解读为右翼；而图 5 所示的画像层级异质性与研究问题2中的意识形态群体并非彼此独立。

B.4 框架、人物画像排序与画像层级

两种筛选框架下均出现归因惩罚，其中在MFT人物画像中，Sánchez

		IW人物画像				MFT人物画像			
主题	等级	画像	L	C	R	概况	L	C	R
移民	high	包容、多元主义	82	16	2	人道主义包容	88	12	0
	mid	中心档案	16	54	30	非极化	22	72	6
	low	限制，传统顺序	0	36	64	限制，边境管控	0	0	100
经济	high	物质保障	0	70	30	功绩与贡献	0	0	100
	mid	中心档案	14	62	24	非极化	22	72	6
	low	自我表达的质量work	74	26	0	社会再分配保障	100	0	0
家庭开支高		可负担性、物质保障	0	44	56	可负担性、基本需求	100	0	0
	mid	中心档案	12	58	30	非极化	22	72	6
	low	自主性，自我表达	74	24	2	市场，个人责任	0	0	100
腐败	high	制度问责	10	72	18	强问责敏感度	36	30	34
	mid	中心档案	16	54	30	中等问责敏感度	22	72	6
	low	个性化权威	20	48	32	较弱的问责敏感度	30	66	4

表19：各主题条件化画像层级中选出的人物画像的意识形态群体：50个人物画像中自我定位为左翼（L，1–4）、中间派（C，5–6）或右翼（R，7–10）的比例（%）。同一人物画像若因多个主题而被选中，则在每个主题中均计入。画像描述摘自表 15。

和 Feijóo 的效应略大（表 20）。这一模式不受每个单元格保留的人物画像数量影响：从排名最佳的人物画像到完整的前50名，所有嵌套队列均产生负向归因效应，且 Abascal 始终受到最大的惩罚。Sánchez 和 Feijóo 的相对顺序在不同 IW 队列间略有变化，但两者的效应仍然接近（表22）。图 5 展示了高、中、低档案等级下平均立场和归因效应的变化幅度。这些等级是按主题划分的价值观画像，并非左右意识形态，因此不能将这种异质性解读为研究问题2的证据。

C 附加结果

C.1 发言者归属

表 21 和图6 展示了四项评分中的说话者归因效应：在每个框架–主题–说话者单元格中，点明领导人身份都会降低一致性、合法性和支持度，并提高感知冒犯性。表 23给出了表4中主题估计值的置信区间。

C.2 意识形态一致性

表格25 至29 分别按群体、框架与群体、框架与主题、主题与群体以及结果给出了研究问题2的估计值；表 24 给出了各位置的效应。一致性、合法性与支持表现出与立场相同的一致性模式，但数值较大，因为它们采用1–5量表；冒犯性则呈现相反的符号，因为负向对比意味着，相较于右翼群体，左翼群体认为A的信息不如B的信息那么冒犯。

C.3 绝对回应水平

归因效应是相较于同一帖子匿名版本的差异。表 30给出了这些效应所依据的各项水平。在未注明作者时，一半回答支持该帖子，四分之一反对该帖子（平均立场为 0.24），因此这些惩罚效应并未受到地板效应或天花板效应的限制。标出姓名后，回答从支持转向反对，中立回答的比例则维持在四分之一左右：以 Abascal 的名义发布时，支持率从 50 % 降至 40%，反对率从 25% 升至 34%。在任何姓名之下，家庭成本帖子仍获得多数支持，包括以 Abascal 的名义发布的情况（54%）。未注明作者时，腐败帖子得到的回答接近中立（-0.04）；而无论标出谁的姓名，多数人都反对这些帖子（54–57%）。

两种框架在水平而非效应上存在更大差异。在所有条件下，MFT人物画像都比IW人物画像更加有利（无作者时为0.39，对比0.10），这一差距大于任何归因效应。因此，这些水平描述的是每种框架的人物画像在同一模型下如何作答。它们并不是对公众支持度的估计

框架	发言者	Δ 立场	95%置信区间
IW	Sánchez	-0.079	[-0.092, -0.066]
IW	Feijóo	-0.081	[-0.093, -0.070]
IW	Abascal	-0.183	[-0.200, -0.166]
MFT	Sánchez	-0.122	[-0.143, -0.101]
MFT	Feijóo	-0.118	[-0.136, -0.100]
MFT	Abascal	-0.198	[-0.220, -0.177]

表20：不同人物选择框架下，所有主题的立场说话者归因效应。

结果	框架	效应范围	Neg.	Pos.
一致性	IW	[-0.365, -0.076]	12/12	0/12
一致性	MFT	[-0.676, -0.031]	12/12	0/12
支持	IW	[-0.495, -0.087]	12/12	0/12
支持	MFT	[-0.585,-0.041]	12/12	0/12
合法性	IW	[-0.334,-0.048]	12/12	0/12
合法性	MFT	[-0.524,-0.024]	12/12	0/12
冒犯性	IW	[0.003, 0.316]	0/12	12/12
冒犯性	MFT	[0.014, 0.433]	0/12	12/12

表21：说话者归因效应对 1–5 评分的影响：各框架在十二个主题—说话者单元中的平均效应范围，以及具有负效应或正效应的单元格数量。

框架	人格队列	Sánchez	Feijóo	Abascal
IW	Best	-0.115 [-0.204, -0.026]	-0.067 [-0.122, -0.012]	-0.178 [-0.263, -0.094]
IW	前5名	-0.094 [-0.132, -0.056]	-0.078 [-0.106, -0.049]	-0.173 [-0.214, -0.132]
IW	前20名	-0.075 [-0.094, -0.057]	-0.084 [-0.101, -0.067]	-0.181 [-0.207, -0.155]
IW	前50名	-0.079 [-0.092, -0.066]	-0.081 [-0.093, -0.070]	-0.183 [-0.200, -0.166]
MFT	Best	-0.124 [-0.234, -0.014]	-0.123 [-0.229, -0.016]	-0.204 [-0.324, -0.084]
MFT	前5名	-0.124 [-0.173, -0.074]	-0.114 [-0.152, -0.076]	-0.173 [-0.214, -0.132]
MFT	前20名	-0.119 [-0.147, -0.092]	-0.116 [-0.140, -0.092]	-0.199 [-0.232, -0.166]
MFT	前50名	-0.122 [-0.143, -0.101]	-0.118 [-0.136, -0.100]	-0.198 [-0.220, -0.177]

表22：嵌套人格群体中的说话者归因效应对立场的影响（每个单元格中排名最佳的人格、排名前 5、排名前 20 以及全部 50 个人格）。单元格报告平均效应 [95% 置信区间]。

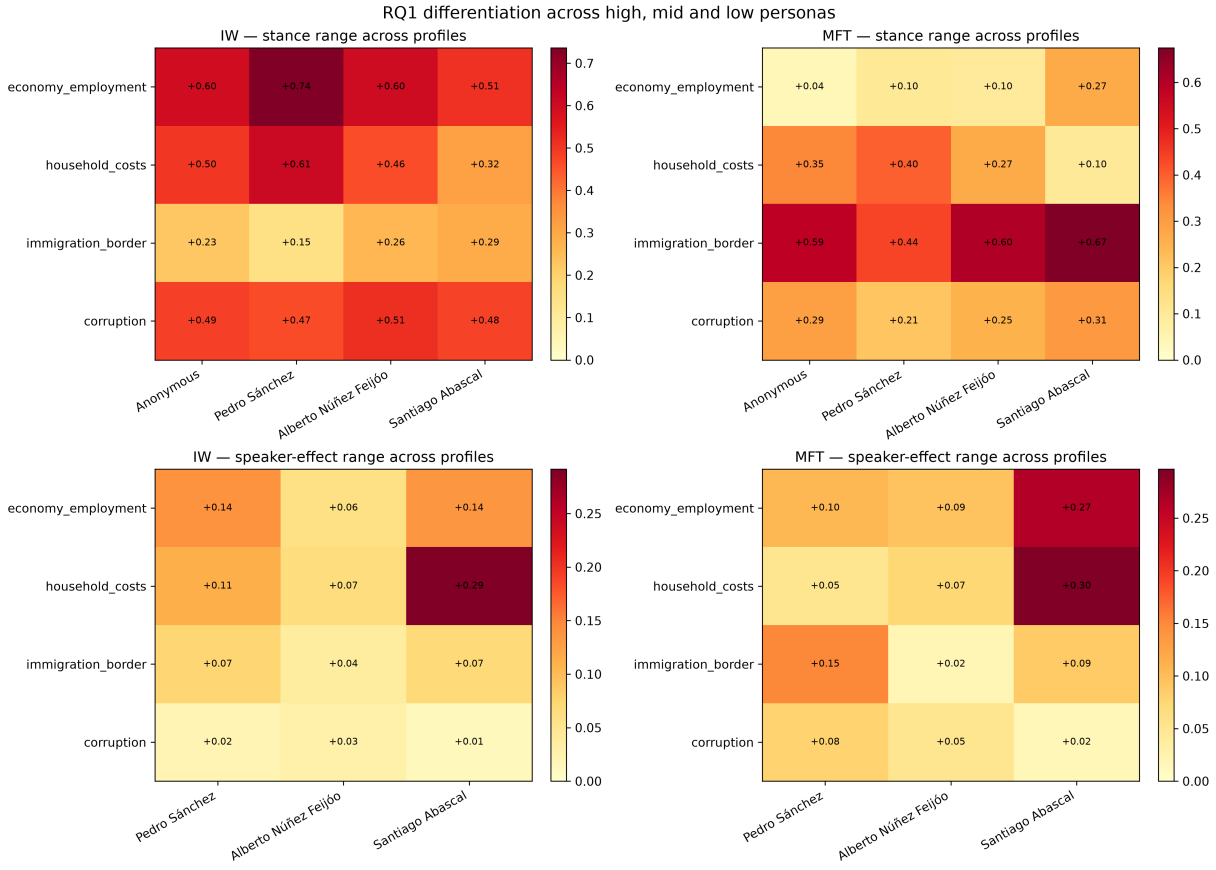

图5：高、中、低档案等级之间的变异。上方面板：对于每个主题和条件，三个等级之间平均立场的范围（最大值减最小值）。下方面板：说话者归因效应的范围。左侧面板为 IW，右侧面板为 MFT。这些等级代表价值档案，而非意识形态。

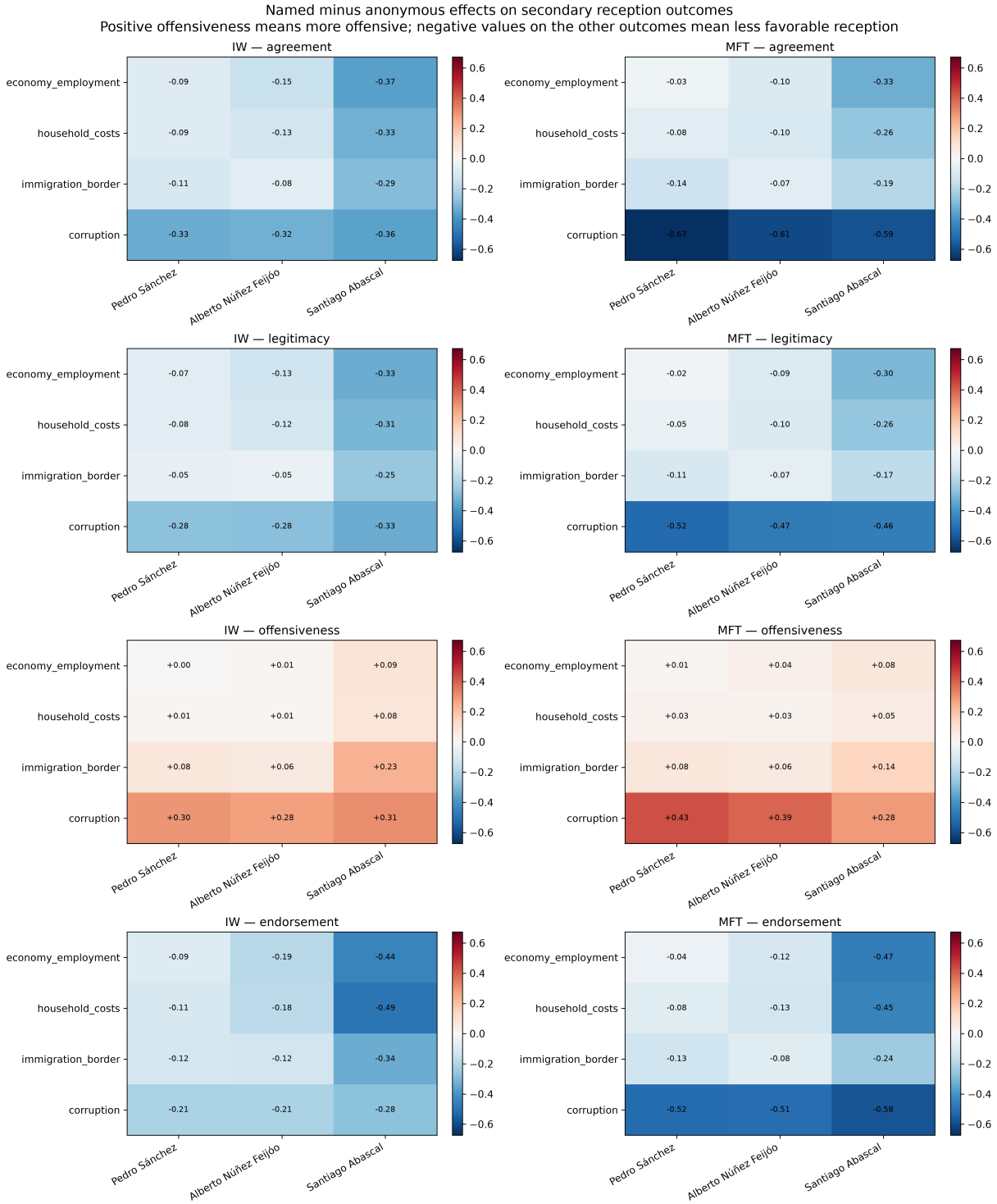

图6：按主题、说话者和框架划分的一致性、合法性、冒犯性和支持方面的说话者归因效应。

框架	主题	Sánchez	Feijóo	Abascal
IW	经济与就业	-0.030 [-0.046, -0.014]	-0.055 [-0.070, -0.040]	-0.186 [-0.210, -0.161]
IW	家庭开支	-0.051 [-0.072, -0.031]	-0.057 [-0.078, -0.036]	-0.168 [-0.207, -0.130]
IW	移民与边境	-0.056 [-0.077, -0.034]	-0.030 [-0.052, -0.007]	-0.164 [-0.196, -0.131]
IW	腐败	-0.194 [-0.219, -0.168]	-0.173 [-0.197, -0.149]	-0.202 [-0.233, -0.170]
MFT	经济与就业	-0.008 [-0.024, 0.008]	-0.042 [-0.060, -0.025]	-0.182 [-0.212, -0.152]
MFT	家庭开支	-0.053 [-0.075, -0.032]	-0.047 [-0.064, -0.029]	-0.132 [-0.165, -0.098]
MFT	移民与边境	-0.065 [-0.090, -0.040]	-0.031 [-0.049, -0.012]	-0.101 [-0.130, -0.072]
MFT	腐败	-0.395 [-0.436, -0.355]	-0.348 [-0.383, -0.312]	-0.335 [-0.376, -0.295]

表23：按主题和框架划分的说话者归因效应对立场的影响。单元格报告平均效应 [95% 置信区间]。

位置	1	2	3	4	5	6	7	8	9	10
人物画像	1	63	129	65	239	47	40	152	41	14
Sánchez	-0.002	0.001	-0.046	-0.097	-0.114	-0.154	-0.124	-0.126	-0.151	-0.140
Feijóo	-0.145	-0.110	-0.112	-0.129	-0.120	-0.125	-0.080	-0.048	0.008	-0.046
Abascal	-0.328	-0.342	-0.270	-0.263	-0.194	-0.198	-0.109	-0.064	-0.037	-0.104

表24：在两种框架和所有主题下，各个诱导的左–右位置对立场的说话者归因效应（图中绘制的数值2a；置信区间见图）。位置 1 仅包含一个人物画像。

人物群体	人物画像	Sánchez	Feijóo	Abascal
左派（1–4）	258	-0.047 [-0.066, -0.027]	-0.116 [-0.133, -0.099]	-0.288 [-0.313, -0.263]
中间派（5–6）	286	-0.123 [-0.141, -0.105]	-0.121 [-0.137, -0.105]	-0.195 [-0.212, -0.178]
右翼（7–10）	247	-0.129 [-0.149, -0.109]	-0.049 [-0.067, -0.030]	-0.072 [-0.090, -0.053]
All	791	-0.100 [-0.114, -0.086]	-0.100 [-0.112, -0.087]	-0.191 [-0.206, -0.175]

表25：不同引导出的意识形态群体在立场上的说话者归因效应，涵盖两种框架和所有主题。单元格报告平均效应 [95%置信区间，按消息和人格进行双向聚类稳健估计]。

框架 群体	人物画像	Sánchez	Feijóo	Abascal
IW	Left	120 -0.023 [-0.043, -0.003]	-0.091 [-0.108, -0.074]	-0.291 [-0.320, -0.261]
IW	中间派	196 -0.081 [-0.097,-0.066]	-0.089 [-0.104,-0.074]	-0.169 [-0.189,-0.150]
IW	右翼	125 -0.134 [-0.158, -0.110]	-0.056 [-0.080, -0.031]	-0.096 [-0.122, -0.070]
MFT	Left	138 -0.063 [-0.091, -0.036]	-0.133 [-0.157, -0.109]	-0.286 [-0.318, -0.254]
MFT	中间派	90 -0.187 [-0.218, -0.156]	-0.170 [-0.196, -0.143]	-0.234 [-0.259, -0.210]
MFT	右翼	122 -0.125 [-0.152, -0.098]	-0.043 [-0.068, -0.018]	-0.052 [-0.075, -0.030]

表 26：不同框架和引导出的意识形态群体下，立场的说话者归因效应，所有主题。单元格报告主题效应均值 [95% 置信区间]。

框架 主题		Sánchez – Feijóo	桑切斯—阿巴斯卡尔	Feijóo – Abascal
Both	所有主题	0.150 [0.128, 0.172]	0.298 [0.264, 0.333]	0.148 [0.124, 0.173]
Both	经济	0.175 [0.143, 0.208]	0.321 [0.274、0.368]	0.146 [0.116, 0.176]
Both	家庭开支	0.143 [0.104、0.182]	0.366 [0.305, 0.427]	0.223 [0.177, 0.268]
Both	移民	0.150 [0.114、0.186]	0.196 [0.143, 0.249]	0.045 [0.003, 0.088]
Both	腐败	0.081 [0.032, 0.130]	0.210 [0.148, 0.272]	0.128 [0.072, 0.184]
IW	所有主题	0.147 [0.118, 0.176]	0.306 [0.261, 0.351]	0.159 [0.123, 0.195]
IW	经济	0.159 [0.118, 0.200]	0.286 [0.229、0.344]	0.127 [0.080, 0.175]
IW	家庭开支	0.170 [0.111, 0.229]	0.426 [0.339, 0.513]	0.256 [0.187, 0.326]
IW	移民	0.143 [0.096, 0.190]	0.165 [0.099, 0.230]	0.022 [-0.038, 0.081]
IW	腐败	0.056 [-0.008, 0.119]	0.205 [0.122, 0.289]	0.150 [0.077, 0.222]
MFT	所有主题	0.152 [0.124, 0.180]	0.295 [0.250、0.339]	0.143 [0.113、0.173]
MFT	经济	0.187 [0.147, 0.226]	0.342 [0.284, 0.400]	0.155 [0.121, 0.189]
MFT	家庭开支	0.123 [0.081、0.166]	0.325 [0.258, 0.391]	0.201 [0.149, 0.253]
MFT	移民	0.157 [0.108、0.206]	0.224 [0.159, 0.289]	0.067 [0.020, 0.115]
MFT	腐败	0.121 [0.044, 0.197]	0.278 [0.200, 0.356]	0.157 [0.071, 0.243]

表 27：按框架和主题划分的立场一致性对比，$(\Delta_A - \Delta_B)_{\text{left}} - (\Delta_A - \Delta_B)_{\text{right}}$。正值表示一致。单元格报告估计值 [95% 置信区间]。

主题	群体	人物画像	Sánchez	Feijóo	Abascal
经济	Left	105	0.040 [0.026、0.055]	-0.077 [-0.098, -0.057]	-0.287 [-0.322, -0.251]
经济	中间派	115	-0.043 [-0.060, -0.026]	-0.053 [-0.069, -0.037]	-0.170 [-0.193, -0.147]
经济	右翼	80	-0.063 [-0.084, -0.042]	-0.005 [-0.025, 0.015]	-0.069 [-0.092, -0.045]
家庭开支	Left	104	-0.019 [-0.040, 0.002]	-0.075 [-0.094, -0.057]	-0.293 [-0.338, -0.249]
家庭开支	中间派	99	-0.047 [-0.069, -0.025]	-0.072 [-0.094, -0.049]	-0.144 [-0.175, -0.113]
家庭开支	右翼	97	-0.094 [-0.121, -0.066]	-0.007 [-0.033, 0.020]	-0.002 [-0.037，0.032]
移民	Left	104	-0.004 [-0.022, 0.013]	-0.043 [-0.060, -0.026]	-0.165 [-0.203, -0.127]
移民	中间派	95	-0.063 [-0.087, -0.040]	-0.044 [-0.068, -0.020]	-0.152 [-0.183, -0.121]
移民	右翼	101	-0.115 [-0.149, -0.082]	-0.004 [-0.033, 0.026]	-0.080 [-0.119, -0.041]
腐败	Left	67	-0.338 [-0.387, -0.290]	-0.329 [-0.376, -0.282]	-0.387 [-0.442, -0.331]
腐败	中间派	171	-0.266 [-0.300，-0.232]	-0.243 [-0.272，-0.214]	-0.260 [-0.292，-0.228]
腐败	右翼	62	-0.326 [-0.372、-0.279]	-0.235 [-0.281、-0.188]	-0.164 [-0.205、-0.123]

表28：两种框架下，按主题和引导出的意识形态群体划分的立场说话者归因效应。单元格报告平均效应 [95% 置信区间]。

结果	Sánchez – Feijóo	桑切斯—阿巴斯卡尔	Feijóo – Abascal
立场	0.150 [0.128, 0.172]	0.298 [0.264，0.333]	0.148 [0.124，0.173]
一致性	0.288 [0.252，0.324]	0.559 [0.499，0.619]	0.271 [0.229、0.314]
合法性	0.258 [0.226, 0.291]	0.530 [0.473, 0.587]	0.271 [0.231，0.312]
支持	0.397 [0.357, 0.437]	0.740 [0.672，0.808]	0.343 [0.293, 0.393]
冒犯性	-0.022 [-0.036，-0.007]	-0.138 [-0.167, -0.109]	-0.117 [-0.143，-0.091]

表29： 按结果划分的一致性对比，涵盖两者框架与所有主题。单元格报告估计值及 [95%置信区间].

帖子。

表 31 按画像层级拆分匿名层级与归因效应。在某一主题内，各层级遵循其价值档案。在MFT人物画像中，“限制与边境管控”层级支持匿名移民帖子（0.51），而“人道主义包容”层级则略微反对这些帖子（-0.07），且“可负担性与基本需求保护”层级对家庭成本帖子最为赞成（0.74）。Abascal在经济方面的惩罚在MFT“再分配与社会保护”层级最大（-0.31），而在“功绩与经济贡献”层级较小（-0.04）。各层级不能跨框架比较：在腐败方面，高问责层级在IW人物画像中最为批判（-0.45），而在MFT人物画像中最为赞成（0.31）。

表 32按发帖领导人对帖子进行划分。在匿名阅读时，Sánchez 的帖子平均立场为 0.47，Feijóo 的为 0.28，Abascal 的为-0.35。在每个意识形态群体和两种框架下，Abascal 的帖子平均立场都最低；即使是右翼人格，对这些帖子也持近乎中立的立场（-0.03，而 Feijóo 的帖子为 0.40）。归因效应将每条帖子与其自身进行比较，因此内容上的差异不会影响归因效应。这描述的是模型如何接收每位领导人的帖子，而非公众如何接收这些帖子。部分原因可能是模型认出了作者，这也是匿名条件并非完全匿名的原因之一（参见“局限性”）。

人物画像	主题	无作者	Sánchez	Feijóo	Abascal
Both	所有主题	0.24 (50/25/25)	0.14 (46/23/31)	0.14 (45/24/31)	0.05 (40/26/34)
	经济与就业	0.36 (54/27/19)	0.34 (54/26/20)	0.31 (52/27/21)	0.17 (44/30/27)
	家庭开支	0.47 (62/22/15)	0.42 (60/21/19)	0.42 (60/23/18)	0.32 (54/24/22)
	移民与边境	0.08 (43/23/35)	0.02 (40/22/38)	0.05 (41/23/36)	-0.05 (34/26/40)
	腐败	-0.04 (35/26/39)	-0.34 (23/20/57)	-0.30 (24/21/54)	-0.31 (24/22/55)
IW	所有主题	0.10 (40/30/30)	0.02 (38/27/36)	0.02 (37/28/35)	-0.08 (31/29/40)
	经济与就业	0.20 (43/33/24)	0.17 (43/31/26)	0.14 (40/33/26)	0.01 (33/35/32)
	家庭开支	0.34 (54/27/20)	0.29 (52/25/23)	0.28 (51/27/22)	0.17 (44/28/27)
	移民与边境	0.00 (37/26/37)	-0.06 (35/25/40)	-0.03 (35/26/39)	-0.17 (27/29/44)
	腐败	-0.21（25/28/46）	-0.40（19/22/59）	-0.38 (20/23/58)	-0.41（18/22/59）
MFT	所有主题	0.39 (59/21/20)	0.27 (54/19/27)	0.27 (53/20/26)	0.19 (48/23/29)
	经济与就业	0.52 (65/21/13)	0.51 (65/21/14)	0.48 (63/21/16)	0.34 (55/25/21)
	家庭开支	0.60（71/18/11）	0.55（69/17/14）	0.55（69/18/13）	0.47（63/20/17）
	移民与边境	0.16 (48/19/32)	0.10 (46/18/36)	0.13 (47/20/34)	0.06 (41/23/35)
	腐败	0.13 (45/23/32)	-0.27 (27/18/54)	-0.22（29/20/51）	-0.21（29/21/50）

表30：按条件划分的绝对回答：平均立场（-1 反对，0 中立，$+1$ 支持），括号内为完整四元组中支持、中立和反对回答的百分比。研究问题1和RQ3的归因效应，是具名列与无作者列之间的差值。按消息和人物画像进行双向聚类后，均值的95%置信区间半宽为0.04至0.10。

		IW人物画像					MFT人物画像				
		无作者		姓名效应			无作者		姓名效应		
主题	等级	平均支持率		S	F	A	平均支持率		S	F	A
经济与就业	High	-0.18	20	-0.11	-0.08	-0.13	0.51	65	-0.06	0.00	-0.04
	Mid	0.36	53	-0.02	-0.02	-0.16	0.54	66	-0.02	-0.04	-0.19
	Low	0.41	57	0.03	-0.07	-0.27	0.50	65	0.05	-0.09	-0.31
家庭开支	High	0.05	37	-0.12	-0.05	-0.06	0.74	82	-0.03	-0.07	-0.27
	Mid	0.41	56	-0.02	-0.03	-0.10	0.66	74	-0.05	-0.07	-0.14
	Low	0.55	67	-0.01	-0.09	-0.35	0.40	57	-0.08	0.00	0.02
移民与边境	High	-0.11	34	-0.01	-0.05	-0.21	-0.07	36	0.00	-0.04	-0.13
	Mid	-0.01	35	-0.07	-0.03	-0.14	0.04	41	-0.05	-0.04	-0.12
	Low	0.11	42	-0.08	-0.01	-0.14	0.51	68	-0.15	-0.02	-0.05
腐败	High	-0.45	10	-0.18	-0.19	-0.20	0.31	59	-0.44	-0.38	-0.33
	Mid	-0.22	24	-0.20	-0.16	-0.20	0.04	39	-0.38	-0.33	-0.33
	Low	0.04	42	-0.20	-0.16	-0.21	0.02	37	-0.36	-0.34	-0.34

表31：各主题条件化画像等级下的绝对回答及其对立场的归因效应（表15），每个单元格包含50个人物画像。无作者：匿名帖子的平均立场和支持性回答百分比。姓名效应：Sánchez（S）、Feijóo（F）或Abascal（A）减去无作者。每个等级在每个主题和每个框架下都代表不同的价值观画像，因此等级只能在同一主题和框架内进行比较。均值的95%置信区间半宽为0.06至0.13，效应的半宽为0.01至0.06。

			显示的条件				无作者，按人物群体划分		
人物画像	作者：	帖子	无作者	Sánchez	Feijóo	Abascal	Left	中间派	右翼
Both	Sánchez	504	0.47	0.45	0.41	0.27	0.67	0.38	0.35
	Feijóo	360	0.28	0.12	0.15	0.08	0.34	0.15	0.40
	Abascal	211	-0.35	-0.56	-0.51	-0.52	-0.47	-0.46	-0.03
IW	Sánchez	504	0.33	0.30	0.27	0.13	0.63	0.23	0.17
	Feijóo	360	0.12	-0.01	0.02	-0.08	0.25	0.00	0.22
	Abascal	211	-0.49	-0.61	-0.58	-0.62	-0.59	-0.57	-0.25
MFT	Sánchez	504	0.61	0.60	0.55	0.42	0.70	0.64	0.48
	Feijóo	360	0.44	0.24	0.29	0.23	0.41	0.36	0.57
	Abascal	211	-0.22	-0.50	-0.44	-0.42	-0.38	-0.33	0.20

表32：按实际撰写帖子的领导人划分的平均立场。显示的条件：匿名帖子，以及以每位领导人的名义发布的同一帖子。无作者，按人物群体划分：自我定位为左派（1–4）、中间派（5–6）或右翼（7–10）的人物画像所看到的匿名帖子。按条件计算的95%置信区间半宽为0.04至0.07，按群体计算的半宽为0.04至0.10。

D CIS比较详情

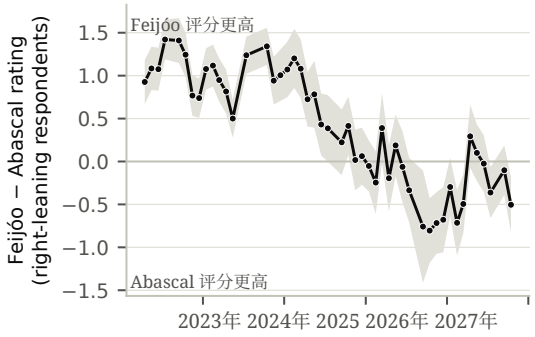

图7：CIS右翼受访者（评分为7–10）对Feijóo和Abascal的平均评分之差，按晴雨表逐一列出，并标示95%置信区间。右翼人格中的排序（Feijóo高于Abascal）与数值高于零的晴雨表一致。

这48份晴雨表涵盖了从2022年4月到2026年10月的每月CIS晴雨表；除8月外，CIS在大选前后，即2023年6月和9月，也未发布晴雨表。领导人评价题是您如何评价他们中的每一位？请从*1*分到*10*分为他们打分，其中*1*分表示您对其评价“非常差”，*10*分表示“非常好”，向受访者询问其所知晓的每位领导人；这是唯一一项在每份晴雨表中都涵盖三位领导人的领导人评价题，因为信任度题项不包括Abascal。意识形态题是人物画像所回答的自我定位问题（附录B.2）。在193,263名受访者中，有8,599人未在量表上进行自我定位，8,900人未评价三位领导人（因为他们不认识其中一人，或回答“不知道”，或未作答）；其余175,764人被纳入分析，因此，领导人的平均值以及领导人之间的差异均来自同一批受访者。评分采用每份晴雨表的全国权重进行加权，自2024年2月起，CIS根据教育情况对该权重进行了调整。在每份晴雨表中，标准误按照加权均值进行线性化处理；合并时则将各份晴雨表视为相互独立的样本。置信区间忽略了抽样设计的聚类效应，因此略显狭窄；比较仅采用估计值的顺序和符号。

表格33列出了受访者和人物画像对领导人的十二项评分差异，表格34展示了十个投放点中各处领导人的排序，而图7 则展示了右翼受访者对Feijóo与Abascal的评分差异，按晴雨表逐一列出。自2025年9月起的8份晴雨表中，其置信区间完全低于零。

领导人 A – B	群体	CIS（评分点）	人物画像（对立场的影响）
Sánchez – Feijóo	Left	3.40 [3.37，3.43]	0.069 [0.055，0.084]
Sánchez – Feijóo	中间派	-1.06 [-1.10，-1.03]	-0.002 [-0.013，0.009]
Sánchez – Feijóo	右翼	-3.50 [-3.54, -3.45]	-0.080 [-0.097, -0.064]
Sánchez – Feijóo	左派 − 右翼	6.90 [6.84，6.95]	0.150 [0.128，0.172]
桑切斯—阿巴斯卡尔	Left	4.83 [4.80，4.85]	0.241 [0.212，0.271]
桑切斯—阿巴斯卡尔	中间派	0.37 [0.34, 0.41]	0.072 [0.053，0.090]
桑切斯—阿巴斯卡尔	右翼	-3.11 [-3.16, -3.05]	-0.057 [-0.079, -0.036]
桑切斯—阿巴斯卡尔	左派 − 右翼	7.93 [7.87、8.00]	0.298 [0.264，0.333]
Feijóo – Abascal	Left	1.43 [1.41，1.44]	0.172 [0.152, 0.192]
Feijóo – Abascal	中间派	1.44 [1.41，1.47]	0.074 [0.059, 0.089]
Feijóo – Abascal	右翼	0.39 [0.34, 0.44]	0.023 [0.007，0.040]
Feijóo – Abascal	左派 − 右翼	1.04 [0.99、1.09]	0.148 [0.124，0.173]

表33：CIS受访者（对48个晴雨表的平均评分进行汇总）和人物画像（$\Delta_A - \Delta_B$ 立场方面，如表3所示）在各群体内领导人之间，以及左派减右翼的差异。正值表示A优于B。单元格报告估计值 [95% 置信区间]。全部十二对的符号均相同。

位置	1	2	3	4	5	6	7	8	9	10
受访者	20,817	12,167	26,512	19,255	40,743	17,161	15,636	11,343	2,632	9,498
人物画像	1	63	129	65	239	47	40	152	41	14
CIS 顺序	S>F>A	S>F>A	S>F>A	S>F>A	F>S>A	F>A>S	F>A>S	F>A>S	A>F>S	A>F>S
人物顺序	S>F>A	S>F>A	S>F>A	S>F>A	F>S>A	F>A>S	F>A>S	F>A>S	F>A>S	
相同顺序	48/48	48/48	48/48	48/48	0/48	9/48	46/48	38/48	26/48	15/48

表34：CIS受访者在各左—右位置上对领导人（S：Sánchez，F：Feijóo，A：Abascal）的排序（汇总平均评分），以及人物画像的排序（说话者归因效应对立场的影响），并列出各位置的受访者人数和人物画像数量。最后一行统计了CIS顺序与人物画像顺序相同的晴雨表数量。在位置9，CIS对Abascal和Feijóo的汇总评分相差0.02分。位置1的人物画像排序仅依据一个人物画像。