Different News

▼ 기획안

[Different News] 임시명

대한민국의 큰 이슈는 대부분 정치적 이슈와 밀접하게 맞닿아 있다.
정치에서 나뉘는 진보(좌파, 사회 전반의 변화나 발전을 추구하는 것을 의미), 보수(우파, 기존의 사회 질서나 전통적인 가치를 유지하는 것을 의미)에 따라 정치에 큰 영향을 끼치는 대한민국 언론사의 성향 또한 구분되어 있다.
한 키워드에 대해(한 이슈에 대해) 자신이 추구하는 방향의 정당과 비슷한 신념의 언론사의 뉴스를 찾아보는 것도 좋지만, **같은 키워드에 대한 이념적 생각을 가진 자들의 뜻도 무엇인지 알아보는 것도 중요하다.**
다른 색의 사람들이 한 공간에 모여 서로의 생각(뉴스)을 읽어보고, 각 생각(뉴스)에 대한 서로의 생각 또한 알 수 있는 웹 사이트 구축이 목표이다.
즉, 목적은 갈라치기가 아닌 같은 생각을 가진 사람의 소리를 듣는 동시에 다른 생각을 가진 사람의 소리 또한 들어보아 **정보의 불균형을 해소**하는 것이다.

▼ 목표 사용자 · 핵심 가치 제안

1. **이슈 비교형 탐독가(정치·시사 관심 높은 2030~4050)**

- JTBD: “한 키워드에 대해 좌·우 기사가 어떻게 다른지 **한 번에 비교**하고 싶다.”
- Pain: 탭을 여러 개 열고 각 언론을 돌아다니는 번거로움, 편향 여부를 빠르게 파악하기 어려움.
- Success: 한 화면에서 대조 포인트가 또렷이 보이고, 신뢰할 수 있는 출처 링크로 바로 이동.

2. **바쁜 직장인/학생(시간 제약, 요약 선호)**

- JTBD: “오늘의 이슈를 **짧게 보고**, 좌·우가 각각 뭘 강조하는지만 **핵심적으로** 알고 싶다.”
- Pain: 길고 산만한 피드, 헤드라인만 보면 왜 다른지 맥락이 안 잡힘.
- Success: 1~3분 내 **요약+대조**로 핵심 이해, 관심 기사만 원문 탐색.

3. **정치·사회 ‘중도/스윙’ 독자(확증편향 탈피 니즈)**

- JTBD: “내 생각과 다른 시각도 **불편하지 않게** 접해 보고 싶다.”
- Pain: 타 진영 기사 노출이 공격적이거나 혐오적 코멘트에 묻힘.
- Success: 중립적 UI/카피, 차분한 비교 구조, 공격적 커뮤니티 억제.

핵심 문제 정의

- 한국의 주요 이슈는 진영별 **프레이밍**이 크게 갈린다.
- 사용자는 한 키워드에 대해 **서로 다른 시각을 빠르게, 공정하게, 안전하게** 비교할 수단이 부족하다.
- 결과적으로 **정보 비대칭**과 **에코챔버**가 강화된다.

핵심 가치 제안 (Value Proposition)

1. **한 화면, 양측 비교**

   한 이슈/키워드에 대해 "좌/우(또는 다양한 축)"의 주요 기사·요약·핵심 문장을 **나란히** 보여줘 대조를 즉시 가능하게 한다.

2. **프레이밍 차이 하이라이트**

   같은 사안을 설명하는 **키워드·인용·수치·주장**의 차이를 자동 추출·표시(예: 강조 문장/용어 차이, 원인·책임 귀인 차이).

3. **공정한 출처 노출·원문 이동**

   자체 기사 게재가 아니라 **제목·리드+메타** 중심(저작권 준수), **원문으로의 명확한 이동**을 최우선.

4. **커뮤니티는 '해석' 중심, 공격 최소화**

   기사/프레이밍 **해석 노트(코멘트)** 중심으로 운영, 인신공격·선동 억제 가드레일 제공.

차별화 포인트

- "좌/우 **대조 경험**"을 일상적인 콘텐츠 소비에 자연스레 녹이는 **UX 중심** 접근.
- **프레이밍 차이**를 '공격'이 아닌 **정보 리터러시**로 번역하는 해석 레이어.
- 저작권·안전성 고려한 **제목/출처/링크** 설계로 리스크 최소화.

핵심 사용자 여정 (Happy Path)

1. 이슈 선택(검색/트렌딩/카테고리) →
2. **양측(또는 다양한 스펙트럼) 기사 카드**가 나란히 등장 →
3. 필요 시 **원문 이동**(새 탭), 혹은 **해석 노트** 열람/작성 →
4. 댓글 작성 및 토론 진행

▼ 용어 정리

- **MVP**(Minimum Viable Product): 제품의 핵심 가치를 제공하기 위한 최소한의 기능들을 정의하는 것

- **TL; DR**: Too Long; Didn't Read의 줄임말로, "너무 길어서 읽지 않았다"는 뜻의 인터넷 속어.
- **JTBD:** Job To Be Done의 줄임말로, 고객이 특정 상황에서 완수하고자 하는 일 또는 해야 할 과업을 의미하며, 이를 해결하기 위해 제품이나 서비스를 '고용'한다는 개념.
- **IA(Information Architecture):** 정보 구조 설계를 의미하며, 서비스나 제품의 콘텐츠를 조직화하고 구조화하여 사용자가 원하는 정보를 쉽게 찾고 이해할 수 있도록 돕는 웹사이트 및 애플리케이션의 설계 작업이다.(메뉴, 화면, 콘텐츠의 계층 구조와 흐름을 시각화, 정보의 배치와 상호작용 등)
- **태깅(Tagging):** 사이트의 관리자가 사이트의 이미지나 텍스트를 관련된 주제나 카테고리의 형태로 형태로 키워트 처리를 해주는 것으로, 쉽게 말하면 태그를 다는 것이다.

▼ 저작권 이슈

- 복제/전송 금지: 언론사 허락 없이 본문, 이미지 등을 복제, 배포, 공중 송신하면 무단 전재에 해당.
  출처만 밝히는 걸로는 합법이 되지 않음.
- 단순 링크(홈 링크): 자유롭게 가능. (여러 개의 언론사 홈페이지 가능)
- 직접 링크(개별 기사 링크): 저작권법 상 '복제, 전송'은 아니라고 보는 판례 경향이 있으나, 업무적, 상업적 이용으로 경제적 이익을 취하면 민법상 부당 이득, 불법 행위로 손해배상 책임이 생길 수 있음. ex) '여러 언론사의 기사를 모아 제공하고 그로 이익을 취하는 경우'가 명시됨.
- 프레임 링크(iframe): 손해 배상 청구가 생길 수 있으므로 지양.
- RSS: 개인 PC 등 개인적 구독 용도만. RSS로 받은 콘텐츠를 공개 배포하거나 재-RSS하면 안 됨.

직접 링크: 특정 웹 페이지, 특히 개별 뉴스나 사진을 직접 링크한 경우, 이용자는 한 개 또는 여러 개의 기사를 그 URL이나 그 기사의 제목을 링크 수단으로 하여 직접 링크 방식으로 이용할 수 있다.
⇒링크 아웃: 기사 URL을 그대로 a태그로 연결하여 클릭하면 그 언론사 뉴스 페이지로 이동(새 탭)
ex) 원문에서 읽기

일반적으로 허용된다(why? 법원의 판단→저작권법 상 '복제'가 아니라 단순 연결이기에.)

but 직접 자체는 저작권 문제는 아니지만, 상업적으로 집계, 트래픽을 모아 **돈을 벌면(영리) 민사상 분쟁 소지.**

⇒제휴, 라이선스를 하면 되지만 **배보다 배꼽이 더 클 수 있다.**

▼ 언론사 별 저작권 특이사항(거의 negative)

- **한겨레, 경향(공통)**
  - https://member.hani.co.kr/help/rules/mypage_help_copyright.hani?type=intellectual_property_protection

- 직접 링크의 적법성 문제와 그로 인해 발생한 경제적 이익의 문제는 별도로 볼 수 있다.
- ⇒ 경제적 이익(영리)을 추구했다면, 민법상 부당 이득, 즉 '법률상 정당한 원인 없이 타인의 재산이나 노동력을 이용해 재산적 이익을 얻고 상대방에게 손실을 준 것'에 해당될 수 있다.

- **오마이뉴스**
  - 일반 사용자도 글을 쓸 수 있어서 저작권이 누구에게 있는지 애매하다. (직접 링크가 되는지 모호함.) ⇒ 잘 명시가 안 되어있다.
- **프레시안**
  - 조합이 생성한 저작물 및 기타 지적 재산권은 조합 법인에 귀속된다.
  - 회원이 사전 승낙 없이 복제, 송신, 출판, 배포, 방송 기타 방법에 의하여 영리목적으로 이용 X
  - ⇒ 조합 법인이 정관 및 규약에 따라 회원에게 귀속된 저작권을 사용할 경우 당 회원의 승낙을 받아야 한다.
- **JTBC(중요)**
  - 회원이 서비스 내에 게시한 게시물 또는 게시물에 포함된 자료 등에 대하여 회사는 게시자의 동의 없이 이를 서비스의 제공 이외의 영리적 목적으로 사용할 수 없음.
  - 회원은 서비스를 이용하여 얻은 정보를 가공, 판매하는 등 서비스를 통해 제공된 자료를 상업적으로 사용할 수 없다.
  - **생성형 AI 학습 데이터 이용 및 데이터 대량 크롤링 제한**
    - 컴퓨터를 이용한 자동화 분석 기술을 통해 추가적인 정보 또는 가치를 생성하기 위한 목적으로 대량의 콘텐트 및 정보를 분석하는 인공지능(AI) 학습 데이터로 이용할 경우 회사와 사전에 반드시 서면 합의해야 합니다.
    - 회사와 합의하지 않은 일체의 콘텐트 무단 수집과 데이터 활용 등의 행위가 데이터 소유권과 콘텐트 및 서비스 저작권 침해에 해당한다고 판단될 경우, 회사는 이에 대한 수집 차단 및 민형사상 책임을 물을 수 있습니다.
    - "ⓒ *JTBC* 무단 이용 (전재 및 재배포, AI 학습 등) 금지" 라고 명시되어 있음
- **조선일보**
  - 이용자는 한개의 기사를 그 URL 또는 그 기사의 제목과 해당 기사 본문의 일부를 함께 표시하는 방법으로 **직접링크할 수 있습니다.** 하지만 이 같은 방식이 **반복적**으로 이뤄질 경우에는 금지되며, **여러 개의 기사**를 그 URL 또는 그 기사의 제목과 해당 기사 본문의 일부를 함께 표시하는 방법으로 **직접링크할 수 없습니다.**
- **중앙일보(중요)**
  - **중앙일보(주)의 사전허가 없이 어떠한 매체에도 직, 간접적으로 변조, 복사, 배포, 출판, 전시, 판매 하거나 상품제작, 인터넷, 모바일 및 데이터베이스를 비롯한 각종 정보서비스 등에 사용하는 것을 금합니다. 특히, 기업이나 기관단체에서 사내 사용을 위해** 자체 데이

터베이스를 구축하거나 이에 해당하는 정보서비스를 하는 것은 사용처가 사내에 한정되고 비영리 목적이라도 저작권법에 위배됩니다.

- 본 서비스의 사용자는 정보를 왜곡, 개작, 변조하지 않을 것을 확약하며 중앙일보는 이를 통해 발생한 문제에 대해서는 책임을 지지 않습니다. 또한 사전의 서면 허가 없이 중앙일보의 정보를 이용하여 영리/비영리 목적의 정보서비스 및 재판매를 할 수 없습니다.

- **동아일보(AI 사용 금지)**
  - 동아일보사와 협의된 목적으로만 사용한다. 저작물을 목적 범위 외에 또는 부당하게 사용한 경우, 신청자는 이로 인해 동아일보사가 입은 손해를 배상하고 법적 책임을 진다.

방안 요약(처음, 지금은 X)

⇒저작권 생각하면 시작도 못함.

1. 미래지향형

   저작권 리스크 최소. 초반 무수익(광고 X) 대신 정보 전달에 중점, 커뮤니티 품질과 파트너십 가능성을 키워 장기 승부

2. 커뮤니티형

   뉴스 원문 다루지 않고 “이슈 체크 및 토론만”으로 빠른 실행, 수익(광고) 가능. 다만 UGC(사용자 생성물) 관리/중재와 평판, 스팸/편파 리스크 관리가 핵심(커뮤니티 집중형)

1) 미래지향형 (수익, 광고 없이 시작. +단순 링크는 활용 가능)

핵심 가설

- “클린하고 논리적인 토론 UX + 저작권 안전 운영”이 신뢰를 만들고, 트래픽이 임계치에 도달하면 언론사와 제휴/라이선스로 윈윈.

제품 핵심

- AI가 핫 이슈 20개 카테고리 자동 선별 → 카테고리 별 토론방 개설(시작 시간 예약 옵션).
- 각 이슈 페이지 구성:
  1. 에디토리얼 요약(자체 작성, 원문 문장 복붙 금지) +이미지는 생성형 AI써서 ?
  2. 출처 링크(언론사로 원클릭 이동)
  3. 논점 카드(찬, 반 포지션, 근거 요구)
  4. 클린 토론 장치(논거 템플릿, 반박/재반박 스레딩, 인신공격 필터)
- 콘텐츠 정책(저작권 안전)

- 원문 직접 인용 최소화/아주 짧게 + 확실한 출처 표시
- 기사 이미지, 표는 사용하지 않음(대신 자체 그래픽/오픈 라이선스 활용)
- RSS는 “링크. 메타”만 사용, 본문 수집/복제 금지
- 사용자 게시물에도 동일 수칙 적용(위반 시 즉시 블록/삭제)

성장/유통

- “논리 배틀 리그” 같은 게임화(근거 점수, 팩트 체크 뱃지).
- 주당 이슈 라이브 토론 이벤트(타임박스, 요약 리포트 생성 → SNS 배포).
- 입소문→임계치 도달 시 제휴/라이선스 협상.

수익(후행)

- 언론사 라이선스 체결 후 합법적 광고/스폰서, 프리미엄 기능(토론 하이라이트, 아카이브, 리서치 대시보드 도입).

주요 리스크 & 완화

- 초반 현금흐름 없음→비용 최소화(서버/모더레이션 툴 우선), 커뮤니티 자율 규범.
- 정치.이념 분쟁→명시적 토론 규칙, 자동.수동 모더레이션 병행, 제재 프로토콜.

핵심 KPI

- 토론 품질 지표(근거 첨부율, 반박/재반박 깊이), 재방문율, 출처 클릭률(언론 유입 기여), 신고 대비 조치시간, 파트너십 관심건수.

2) 커뮤니티형 (뉴스 저작권 논외, “토론만” 운영 + 광고)

핵심 가설

- 원문을 다루지 않고 이슈 토론 그 자체에 집중하면 실행, 수익을 빠르게 올릴 수 있다.

제품 핵심

- AI 선정 이슈 20개 + 토론방.
- 콘텐츠는 전부 UGC(기사 링크 공유는 OK, 본문 복붙 NO).
- “논증력/팩트 레벨” 뱃지, 신고, 정정 흐름.

수익

- 즉시 디스플레이/네이티브 광고 탑재 가능(페이지 RPM 최적화).
- 트래픽 성장 시 제휴/라이선스로 스폰서형 토론, 프로 커뮤니티.

주요 리스크 & 완화

- UGC 저작권/명예훼손/허위정보: 이용약관, 커뮤니티 가이드, 신고-심사-삭제 루프, 권리침해 통지/처리 절차 명시

- 품질 저하, 편파/악성 이용자: 신뢰점수, 느린모드, 초보자 승인제, 전문가/모더 권한.

핵심 KPI

- 활성 작성자/독자 비율, 신고 처리시간, 광고 RPM/세션당 페이지뷰, 장기 유지율.
- +게시글 기능 추가.

==========

▼ IA(Information Architecture)

시작하기 전, 검색창이 필요한가를 고민. 처음엔 필요없다고 생각합니다.

1. 홈

- 상단: 사이트이름/간단한 홈페이지 소개/로그인, 마이페이지
- 본문: “인기 토픽 top 10” 카드 그리드(토픽명, 간단 설명/핵심 키워드 3개, 생성 시각, hot, **썸네일** 등 설명)
- 푸터: 약관, 운영원칙, 신고/문의 등

2. 각 토픽 페이지

- 왼쪽 사이드: 인기 토픽 10개 리스트업(볼륨, 출처 다양성, 좌/우 균형, 조회수 등으로 가중치 부여 정렬)
- 오른쪽 사이드: 참고 언론사 리스트업 or 광고 (더 생각)
- 헤더: 토픽명, 간단한 설명(중도의 입장), 생성 시각

▼ 예시

- 본문: 좌/우 2열 레이아웃(PC), 모바일은 탭 전환(좌/우)

    - 각 열: 언론사 구분 섹션→기사 카드 리스트(원제목, 게재일-언론사명, 바로가기 링크)
    - 각 언론사 별or최신순 페이지네이션 또는 ~~무한스크롤~~
- 하단: 참여 영역의 댓글(라이트)
    - 짧은 의견, 추천/비추천 기능, 정렬(최신/추천순), 신고
    - 추가 기능⇒성향 커밍아웃 라벨, 클린 뱃찌 등

3. 토론방 시작 페이지(토론 시작 페이지로 넘어갈 수 있는 trigger를 만들어야 함.)

- 왼쪽 사이드: 인기 토픽 10개 리스트업
- 오른쪽 사이드: 참고 언론사 리스트업 or 광고 (더 생각)
- 헤더: 간단한 설명(토론 및 토론방 설명)
- 본문: 토론방 생성
    - 토론 주제(선택 or 생성), 시작 시각, 참여자 인원 제한
    - 리스트업 된 hot 토픽에 대한 예시 토론 주제 3~5개 중 선택 가능

4. 토론방(아직 미흡+다른 분들의 의견이 더 좋을 거 같음))

- 오른쪽 사이드: 참여자 이름 리스트업
- 본문: 채팅
- 하단: 관전자(챗 x. but 엄지 등 이모지?로 표현)
    - 규칙 설명란(ex. 인사로 시작 끝, 비방 및 욕설 금지 등)

        ▼ 토론의 주요 원칙

        - **추정의 원칙:** 특정 주장이 증명되지 않은 한, 그 주장이 참이라고 단정해서는 안 됩니다.
        - **평등의 원칙:** 모든 참여자에게 동등한 발언 기회가 주어져야 합니다.
        - **상호 존중의 원칙:** 상대방을 존중하고, 강압이나 협박보다는 설득으로 문제를 해결해야 합니다.
        - **결과 승복의 원칙:** 토론 결과에 대해 승복하고, 토론을 통해 성찰하는 자세가 필요합니다.
- 토론 룰 ex)
    - 토론 시작 시 토론자들은 모두 “매너있는 토론을 진행하겠습니다” 치고 시작하게 하기
    - 자신의 주장 띄워놓기(옆 자신의 이름 아래)

5. Q&A / 메일: ~기사 다뤄주세요, 채팅창 점검 해주세요 등
5. 게시물(마이너 커뮤니티가 될 수도 있음)

▼ 한 기사 예시 구조

- 출처: 파비콘+언론사 명(파비콘 URL 캐싱 필요)
- 제목:기사의 원제목(원문 바로가기: 직접링크)
- ~~요약(1~3줄): 사실 요약 + 짤막한 코멘트(왜 읽을 가치가 있는지)~~ (저작권)
- ~~키워드: 3-5개~~ (저작권)
- 게재일(ex 23분 전, 1시간 전, 어제 / **9월 16일 17:00**)

▼ 예시(기사)

+관련 콘텐츠 누르면 각 성향 언론사들 홈페이지 리스트업(모달?)

+시각화가 중요하다고 생각. 썸네일 필요(뉴스토어에서 사진만 구매 가능(보도 사진 섹션)-추후 논의)

▼ AI 썸네일 가이드라인

- 안전 수칙(핵심)
  - **사람 얼굴, 실존 인물 금지**: 정치인, 연예인, 기자 등 식별 가능한 인물을 닮게 그리거나 합성하면 초상권/명예훼손 리스크가 커짐. 가능하면 아이콘, 추상 일러스트로 대체.
  - **로고, 브랜드, 언론사 마크 금지:** 썸네일에 JTBC, KBS, 삼성 로고 같은 제 3자 상표/마크를 넣지 않기. (혼동, 상표 리스크)
  - **사실 오인 유발 금지:** 범죄/사고 기사에 특정 회사/기관을 연상시키는 건물, 유니폼, 색채/패턴을 쓰지 않기. 묘사는 중립적, 추상적으로.
  - **원본 자동 프리뷰 차단:** 링크의 OG 이미지/문구가 자동으로 뜨지 않게 하고, og:image는 내가 만든 썸네일로 명시.
  - **표시 문구:** 페이지 어딘가에 작게 "썸네일은 AI 생성 일러스트입니다." 정도의 고지(특히 논쟁 이슈일수록 권장됨.)
  - **저작권, 약관 체크:** 사용하는 이미지 생성 모델의 이용약관에서 "출력물의 사용권"을 확인

    - **이의제기 대응:** 문의/클레임 시 신속히 교체, 삭제하는 절차를 준비.
- 실무 팁
    - **프롬프트 가드레일:**
      "*Minimalist illustration of [주제],* ***no faces, no people, no logos, no brands, no text****, abstract icons, neutral colors, 1200×630*"
    - **일관된 디자인 시스템:** 카테고리별 컬러/아이콘 세트(경제=파랑, 정치=보라…)를 정해 혼동 없는 '우리만의' 시각 언어 만들기.
    - **접근성/메타:** 1200×630(또는 1200×628) 권장, alt 텍스트 "(일러스트) [주제] 관련 추상 아이콘" 같이 명확히.
    - **로그 보관:** 생성 프롬프트·시드·버전을 저장(사후 분쟁 시 설명 가능).

▼ 참고 사이트

- **오마이뉴스/논쟁**
    - https://www.ohmynews.com/NWS_Web/Debate/index.aspx
    -
- 한국경제/생글생글
    - https://sgsg.hankyung.com/series/1509001006
    - 활성화 X

토픽 선정 파이프라인(중요)

단계 & 규칙

1. **수집**(크롤러/**RSS**/검색트렌드)
    - RSS 폴링(5분 간격) → `articles` UPSERT(키: `domain + normalized_title + published_at`)
    - 시간 표준: **KST**, 상대·절대 시각 모두 저장(`rel`, `abs`)
2. **임베딩(제목) → 군집화 (AI 허용 소스만)**
    - `article.ai_allowed=TRUE`(AI 허용)인 것만 **제목 임베딩** 생성 → 최근 **24~48h**만 대상으로 HDBSCAN/BERTopic
    - **가드레일**
        - 최소 클러스터 크기: **≥3개 기사**
        - 유사도 평균(클러스터 내) **≥0.80** 미만이면 **보류 큐**

    - 동일 도메인 비중 **>60%**면 “과도 단일 출처” 경고

1. **군집명 자동 생성(일반 주제명)**
    - **출처 문장 복붙 없이** 일반 주제명 생성(예: “정년연장 논의”, “민생쿠폰 지급”)
    - 후보 3개 자동 제안 → **운영자 선택** (선택 전에도 임시명으로 표시 가능)
2. **좌/우 대표 기사 최소 3건 보장**
    - 각 토픽에 대해 좌/우(매체 성향 테이블로 결정) **각 ≥3건** 충족 시 자동 “준비완료”
    - **미충족 시 보강 절차**
        - AI 허용 기사 풀에서 동일 키워드/엔티티로 재검색(최근 72h까지 확장)
        - 그래도 부족하면 **운영자에게 알림**: No-AI 기사 **수동 연결** 요청(좌/우 불균형 배지)
3. **중복·낚시성·오보 리스트 필터**
    - **블랙리스트 도메인** 차단(오보·낚시성 기록된 매체)
    - 제목 패턴(“충격/헉/ㄷㄷ/클릭”) 차감 또는 제외
    - **동일 URL/제목 중복 제거**(UTM 제거, canonical 정규화)
4. **운영자 최종 승인**
    - 토픽 상태: `draft → review → published`
    - 화면에서 확인: 토픽명 / 좌·우 기사 수 / 불균형 배지 / 블랙리스트 제거 내역
    - [병합] 후보 / [분할] 미리보기 제공(자동 제안은 내부만 사용, 숫자 노출 X)

> 표현 원칙(사용자 화면): 원제목·출처·원문링크·게재시각 **만** 노출. (요약·재제목·키워드 없음)

아키텍처(YES-AI) – 구현 디테일

컴포넌트

- **Ingest(NestJS)**: RSS 폴링 → `articles` UPSERT
- **Embed Worker**: `ai_allowed=TRUE` 만 제목 임베딩 → `article_vectors`
- **Clustering Batch**: 15분 단위 → `topic_candidates`, `topic_articles(AUTO)`
- **Scoring Batch**: 토픽 점수(볼륨·출처다양성·좌우균형) → `topic_scores`
- **Admin(React+Nest)**: 후보 검수/병합/분할/No-AI **수동 연결(MANUAL)** → Publish
- **Public API**: `/topics`, `/topics/:id` (좌/우 기사 목록 반환)

최종 FLOW

1. **수집**: RSS → 제목/링크/매체/시간 저장(중복 제거)
2. **토픽 생성**: 허용 매체만 임베딩/규칙으로 묶어 토픽 만들기
   - 수집(RSS) → 제목 임베딩(L2)
   - **DBSCAN 군집** → 작은/품질 낮은 묶음 제거
   - 기존 토픽에 붙이거나 새 토픽 생성
   - **좌/우 최소 보장 체크** → 부족 시 알림
3. **좌/우 라벨**: 매체 성향 테이블로 기사에 좌/우 부여
4. **No-AI 기사 처리**: Admin에서 **드래그&드롭/선택**으로 No-AI 기사를 **특정 토픽에 직접 연결**
5. **검수/게시**: 운영자가 병합·분할·게시 승인
6. **사용자 화면**: 이슈 상세에서 **좌/우 칼럼**에 기사 카드(파비콘+매체명 / 원제목(원문링크) / 게재 시각)

▼ 언론사 RSS 주소

- 경향 신문(제목 + 2~3 문장)
  - https://www.khan.co.kr/help/help_rss.html
- 한겨레(AI 학습 및 활용 금지, 매일 08:59 업데이트)
  - https://www.hanion.co.kr/rssIndex.html
- 오마이뉴스
  - https://rss.ohmynews.com/rss/politics.xml
- 조선일보
  - https://rssplus.chosun.com
- 중앙일보(우회 루트)
  - 중앙일보는 자체 RSS서비스를 중단하여, Google News에서 발췌해서 RSS 긁어올 수 있음.
  - 검색창에 ,로 키워드 입력 후 .com/뒤에 rss붙이기 ⇒ .com/rss/search?q\...
  - (예시) https://news.google.com/rss/search?q=중앙일보%2C 검찰개혁&hl=ko&gl=KR&ceid=KR%3Ako
- 동아일보(AI학습 이용 금지)
  - https://rss.donga.com/

⇒Google News에서 인기 토픽에 대한 각 언론사의 RSS를 가져와서 AI돌려서 관리자(admin)에 보여준 후 관리자가 수동으로 선택한 뒤 사용자(api)에 리스트업.

+정치를 넘어 경제, 사회/국제까지 확장 가능.

▼ JSON 로직

- RSS URL 저장(예시)


```
{
  "sources": [
    {
      "name": "경향신문",
      "domain": "khan.co.kr",
      "side": "LEFT",
      "feeds": [
        "https://www.khan.co.kr/rss/rssdata/politic_news.xml"
      ],
      "allow_ai_processing": true
    },
    {
      "name": "중앙일보",
      "domain": "joongang.co.kr",
      "side": "RIGHT",
      "feeds": ["<여기에 찾은 RSS URL>"],
      "allow_ai_processing": false
    }
  ]
}
```


- 표시용 JSON (사이드 카드 렌더링용)


```
{
  "category": "검찰개혁",
  "updated_at": "2025-09-16T17:00:00+09:00",
  "items": [
  {
    "source": "경향신문",
    "source_domain": "khan.co.kr",
    "side": "LEFT",
    "title": "예시) 수사권·기소권 분리 논쟁 재점화…여야 공방 가열",
    "url": "https://www.khan.co.kr/article/2025xxxxxxx/",
    "published_at": "2025-09-16T10:30:00+09:00",
    "summary": null,
    "image_url": null
    }
```

```
  ]
}
```

- DB 삽입용 JSON (메타데이터 전용)

```
[
  {
  "category": "검찰개혁",
  "source": "경향신문",
  "source_domain": "khan.co.kr",
  "source_side": "LEFT",
  "origin_language": "ko",
  "section": "정치",
  "author": "홍길동",
  "content_type": "ARTICLE",              // ARTICLE | OPINION | EDITORIAL ...
  "url": "https://www.khan.co.kr/article/2025xxxxxxx/",
  "url_canonical": "https://www.khan.co.kr/article/2025xxxxxxx/",
  "published_at": "2025-09-16T10:30:00+09:00",
  "retrieved_at": "2025-09-16T10:45:12+09:00",
  "paywall": false,

  "ai_processing_level": "EMBED_SUMMARIZE", // NONE | EMBED | SUMMARIZE | EMBED_SUMMARIZE
  "license_status": "UNKNOWN",            // OK | BLOCK | UNKNOWN
  "rss_guid": null,
  "title_text": "예시) 수사권·기소권 분리 논쟁 재점화...여야 공방 가열",
  "title_ai": null,                       // 재제목 사용 시
  "summary_ai": null,                     // 자체 요약 저장 시
  "keywords": ["검찰개혁", "수사권-기소권 분리", "검경수사권"],
  "og_image_url": null,                   // 썸네일 수집 시
  "url_hash": "sha256:...",               // 선택
  "title_hash": "sha256:...",             // 선택
  "dedup_key": "sha256:khan.co.kr__수사권·기소권 분리 논쟁 재점화...여야 공방 가열__2025-09-16"
  }
]
```

▼ MVP 구현 순서

v0.1 (주요 화면 + 수동 연결로 가동)

- 기능: 토픽 생성/게시, 좌/우 기사 카드 렌더, No-AI 수동 연결, 매체 성향 관리
- DoD:
  - 기사 카드: 파비콘·매체·원제목(원문 링크)·게재시각(KST)
  - 좌/우 정렬(최신순), 빈 칼럼 안내 문구
  - Admin: 미연결 풀에서 **MANUAL** 연결, 감사로그 저장

v0.2 (AI 허용 자동 연결 + 품질 가드)

- 기능: 임베딩/군집화 배치, 자동 연결(AUTO), “최소 3건 보장” 로직, 병합/분할 제안
- DoD:
  - `ai_allowed=TRUE`만 자동 군집/연결
  - 좌/우 **각 ≥3건** 충족 못하면 **불균형 배지 + 운영자 알림**
  - 병합/분할 실행 → 화면 즉시 반영

v0.3 (운영 편의 & 안정화)

- 기능: 블랙리스트/낚시성 필터, 검색트렌드 기반 **빈 토픽 시드**(이름만), 클릭 로깅, 간단 통계
- DoD:
  - 블랙리스트 반영 로그, 제외 사유 표시
  - 트렌드 키워드로 빈 토픽 생성(운영자가 수동 연결해 채움)
  - 클릭/체류 로그 수집(내부 대시보드 초간단 표)

Node(API·Admin·크롤러) + **Python(임베딩/HDBSCAN)** 분리 마이크로서비스. 통신은 HTTP or 큐, 저장은 공용 DB/벡터DB

- **제목 임베딩** (Sentence-Transformers 등)
- **L2 정규화** (코사인≈유클리드 되게)
- **DBSCAN 실행**
  - `metric='euclidean'` (정규화했으니 코사인과 동치)
  - `min_samples=2 ~ 3` (핵심 포인트 최소 이웃 수)
  - `eps`는 **거리 임계값** → 아래 튜닝법 참고
- **후처리**
  - 노이즈 라벨(-1) 제외
  - 클러스터 크기 < 3이면 버리기(우리 정책: 토픽 최소 3개)

- 각 클러스터 **센트로이드(평균 벡터)** 계산 → 기존 토픽과 코사인 유사도 ≥ **0.86**이면 합류, 아니면 새 토픽
- 좌/우(매체 성향) **각 ≥3건** 보장 못하면 운영자에게 **수동 연결** 알림