# 대항선전(對抗宣傳, Counter-Propaganda)

## 이론·역사·전략·심리·정보전·민주적 통제와 정보생태계 회복탄력성


**작성자: 코리아베스트** https://koreabest.org

**작성자: The American Newspaper** https://americannewspaper.org

2026 년 8 월 12 일

**문서 성격** 본 보고서는 연구·교육 목적의 방어적 분석이다. 특정 정치집단이나 인구집단을 조작하기 위한 표적 설득 지침이 아니라, 선전·허위정보를 탐지·검증하고 민주적 정보환경의 신뢰와 회복탄력성을 강화하는 원리와 사례를 다룬다.

# 요약: 대항선전의 핵심 명제

대항선전은 상대방의 선전 메시지 하나를 “반박”하는 기술이 아니라, 상대의 정보작전이 작동하는 전체 사슬 - 정보원, 메시지, 프레임, 서사, 매체, 네트워크, 알고리즘, 청중의 사전신념, 그리고 제도적 신뢰 - 을 분석하고 그 영향력을 약화하거나 무력화하는 방어적 커뮤니케이션 활동의 집합이다. 미 합동교범 계열의 오래된 정의가 counterpropaganda 를 “선전을 무력화하거나 그 효과를 완화하기 위한 산출물과 행동”으로 넓게 정의한 것도 이 때문이다.[41] 현대적 의미에서는 사실검증, 출처 폭로, 프레임 해체, 대안서사, 프리벙킹, 플랫폼 조치, 미디어 리터러시, 독립언론 지원, 제도 신뢰 회복까지 포함하는 다층적 개념으로 이해하는 것이 적절하다.

- 선전은 반드시 거짓일 필요가 없다. 사실의 선택·배열·강조, 감정 자극, 반복, 정체성 결합, 출처 위장만으로도 선전 효과가 발생할 수 있다. 따라서 “거짓말을 사실로 교정하면 끝”이라는 접근은 불충분하다.[1][2]
- 대항선전의 승패는 메시지의 논리적 정확성보다 “누가, 언제, 어떤 청중에게, 어떤 채널과 사회적 관계 속에서 말하는가”에 크게 좌우된다. 신뢰받지 못하는 정보원이 정확한 사실을 전달해도 실패할 수 있다.[6][15]
- backfire effect 는 가능한 위험이지만 보편적 법칙으로 보기는 어렵다. 최근 연구는 교정이 평균적으로 오정보 믿음을 낮추는 경우가 많으며, 역효과는 특정 조건에서 발생하는 이질적 현상으로 보는 쪽에 무게를 둔다.[17]
- 프리벙킹·심리적 접종은 조작기법을 미리 보여주고 약한 반론과 반박을 경험하게 하여 인지적 저항력을 키운다. 메타분석에서는 허위정보 신뢰도를 낮추고 진짜 정보의 신뢰도를 높이는 효과가 확인되었지만, 공유행동까지 항상 바꾸는 것은 아니다.[16][18]
- 디지털 시대의 핵심 위협은 “콘텐츠”보다 “배포 인프라”다. 봇·트롤·익명계정·광고·추천 알고리즘·커뮤니티 허브가 사회적 증거와 반복노출을 인위적으로 만들 수 있다. 생성형 AI 는 이 구조의 속도·규모·다양성을 증폭한다.[11][22][23]
- 민주주의의 대항선전은 검열과 구별되어야 한다. 사실성, 출처 공개, 독립적 검증, 비례성, 법적 통제, 언론 자유, 시민의 선택권, 오류 정정 가능성이 핵심 경계다. 국제인권법상 표현의 자유 제한은 법률성·정당한 목적·필요성·비례성의 엄격한 심사를 받는다.[30]
- 최종 목표는 상대방보다 “더 강한 선전”을 생산하는 것이 아니라, 거짓말이 덜 이익이 되고 오류가 빨리 교정되며 신뢰할 만한 정보가 지속적으로 공급되는 정보생태계를 만드는 것이다.[20][21]

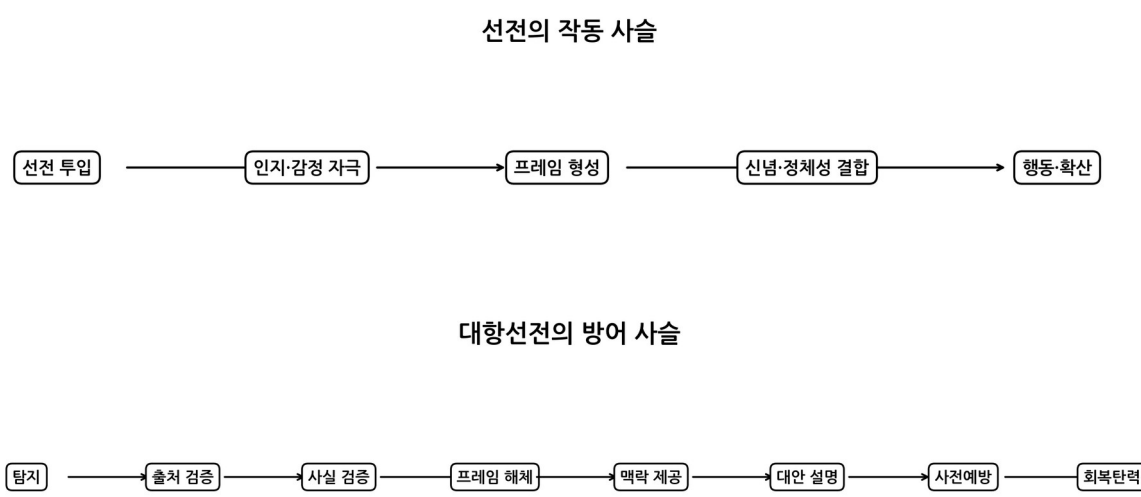


그림 1. 선전의 작동 사슬과 대항선전의 방어 사슬을 결합한 통합 모델

# 목차

# 1. 정의·어원·개념적 기원

## 1.1 정의: "반박"보다 넓은 개념

대항선전(counter-propaganda)은 특정 행위자가 유통하는 선전의 신뢰성, 설득력, 확산력 또는 행동 유도 효과를 약화·중화·무력화하기 위해 수행하는 분석·커뮤니케이션·교육·플랫폼·제도적 대응의 총칭이다. 고전적 군사문헌에서는 이를 상대 선전의 효과를 "nullify or mitigate"하는 행동으로 정의했고,[41] Linebarger 는 적 선전의 분석 자체가 정보 획득과 counterpropaganda operation 의 기초가 된다고 보았다.[10] 현대 연구에서는 단순한 "같은 채널에서 반대 주장을 방송하는 것"을 넘어, 신뢰할 수 있는 대안 정보공급, 출처 투명성, 사회적 회복탄력성, 독립언론 및 미디어 리터러시까지 연결한다.[14][20][21]

**핵심 구분** 대항선전은 "거짓에 대한 거짓"이 아니다. 방어자가 사실성·투명성·검증 가능성을 포기하면 단기적 반박에서 이겨도 장기 신뢰자본을 훼손하여 전략적으로 패할 수 있다.

## 1.2 한자 對抗宣傳의 의미

| 글자 | 기본 의미 | 개념적 기능 |
|---|---|---|
| 對(대) | 마주하다, 상대하다, 대응하다 | 상대의 메시지·행위·영향에 맞서는 관계성을 표시 |
| 抗(항) | 저항하다, 버티다, 거스르다 | 단순한 반응이 아니라 영향력을 억제·중화하는 방어 의도를 표시 |
| 宣(선) | 널리 알리다, 드러내다, 선포하다 | 공개적 커뮤니케이션·전달의 적극성을 표시 |
| 傳(전) | 전하다, 퍼뜨리다, 계승하다 | 메시지의 확산·전파·매체적 성격을 표시 |

따라서 對抗宣傳은 문자 그대로 "상대의 선전·전파에 맞서 저항하는 선전/커뮤니케이션"이라는 조합이다. 다만 이 표현은 고전 한문에서 정립된 정치학 용어라기보다 20 세기 대중정치·전쟁·심리전의 발전과 함께 번역·정착한 현대 복합어로 보는 것이 타당하다.

## 1.3 영어 counter-propaganda 의 형성과 사용

영어권에서는 counter-propaganda 와 counterpropaganda 가 병용된다. "counter-"는 반대·대응·상쇄를 뜻하며 propaganda 에 결합하여 상대 선전의 효과를 막는 활동을 가리킨다. 접근 가능한 문헌만으로 최초 사용자를 확정하기는 어렵지만, 제 1 차 세계대전 이후 선전 연구가 제도화되고 제 2 차 세계대전의 심리전이 전문화되는 과정에서 anti-propaganda, rebuttal, counter-propaganda 같은 표현이 군사·정보·언론 실무에서 널리 등장한다. 1948 년 Paul M. A. Linebarger 의『Psychological Warfare』는 "counterpropaganda operations"를 명시적으로 논의한다.[10] 이후 미군 교범과 정보작전 문헌에서도 counterpropaganda 가 독립 기능 또는 보조 기능으로 사용되어 왔다.[41]

중요한 점은 이 용어가 단일 학문분과의 엄밀한 고정개념이라기보다 군사심리전, 공공외교, 전략커뮤니케이션, 정보작전, 정치커뮤니케이션, 허위정보 대응이 겹치는 "실천적 중간개념"이라는 것이다. 따라서 연구에서는 먼저 어떤 기관이 어떤 의미로 사용하는지 정의를 확인해야 한다.

# 2. 인접 개념과의 경계

대항선전을 이해하려면 무엇과 같은지보다 무엇과 다른지를 먼저 구분해야 한다. 아래 표에서 "대항선전"은 상대 선전의 영향 무력화가 1 차 목적이고, 다른 개념들은 더 넓거나 더 좁은 목적을 가진다.

| 개념 | 1 차 목적 | 주요 행위자 | 대항선전과의 차이 |
|---|---|---|---|
| Propaganda | 인식·감정·행동을 특정 방향으로 유도 | 국가·정당·운동·기업 등 | 사실·선택적 사실·거짓 모두 가능; 체계적 설득이 핵심 |
| Counter-propaganda | 상대 선전의 효과를 약화·중화 | 정부·군·언론·시민사회·플랫폼 | 반박뿐 아니라 출처·프레임·네트워크· |

| | | | |
|---|---|---|---|
| | | | 회복탄력성까지 포함 |
| Counter-messaging | 특정 메시지에 대응하는 메시지 제공 | 정부·NGO·캠페인 | 대항선전보다 메시지 단위에 집중 |
| Strategic communication | 목표·행동·메시지를 통합하여 장기적 효과 창출 | 정부·군·조직 | 대항선전을 포함할 수 있으나 전체 전략과 행동의 일관성이 핵심 |
| Public diplomacy | 외국 대중과 장기 관계·이해·신뢰 형성 | 외교기관·문화기관 | 상호작용·교류가 중요; 적대 선전 대응만이 목적이 아님 |
| PSYOP/MISO | 승인된 목표청중의 인식·태도·행동에 영향 | 군·안보조직 | 작전지원 목적; 정보환경 분석과 표적청중 분석 포함 |
| Information operations | 정보 관련 능력을 통합·조정하여 작전 효과 달성 | 군·안보조직 | 사이버·기만·전자전 등 비메시지 수단도 결합 가능 |
| Information warfare | 정보 우위·교란·보호를 둘러싼 경쟁 | 국가·군·비국가행위자 | 공격·방어, 기술·심리·네트워크를 포괄하는 넓은 경쟁개념 |
| Influence operations | 인식·의사결정·행동에 영향 | 국가·조직·네트워크 | 은밀·공개 수단 모두 가능; 정당성·윤리 문제가 큼 |
| Cognitive warfare | 인간의 인지·판단·사회적 의사결정 자체를 경쟁공간으로 봄 | 군사·안보·연구기관 | NATO 논의에서도 합의된 단일 법적 정의라기보다 발전 중인 개념.[24][25] |
| Fact-checking / debunking | 구체적 사실주장의 참·거짓 검증·교정 | 언론·검증기관·학계 | 가장 좁은 층위; 서사·정체성·배포망 전체는 다루지 못할 수 있음 |
| Pre-bunking / inoculation | 오정보 접촉 전에 조작기법에 대한 저항력 형성 | 교육·플랫폼·공공기관 | 사후 반박이 아니라 예방적 인지면역 형성.[16][18] |
| Media literacy | 정보의 출처·의도·증거를 평가하는 시민 역량 | 교육·언론·시민사회 | 장기 회복탄력성 전략 |
| Counter-disinformation | 고의적 허위정보 캠페인의 탐지·노출·완화 | 정부·언론·플랫폼·시민사회 | “허위”와 “고의성”에 초점; 선전은 사실기반일 수도 있어 대항선전보다 범위가 다름 |

**개념적 함정** “선전=거짓말, 대항선전=팩트체크”라는 이분법은 틀리다. 선전은 진실한 사실을 선택적으로 배열할 수 있고, 팩트체크가 사실오류를 교정해도 감정·정체성·프레임·네트워크 효과가 남을 수 있다.

# 3. 핵심 목적과 전략 원리

## 3.1 상대 선전의 신뢰성 약화

대항선전의 가장 직접적인 목적은 특정 주장 하나를 논파하는 것이 아니라 “이 정보원과 이 배포망을 앞으로도 믿어도 되는가?”라는 장기 판단을 바꾸는 것이다. 반복적으로 확인 가능한 오류, 조작된 이미지, 출처 위장, 서로 모순되는 버전, 행동과 주장 사이의 불일치를 투명하게 제시하면 정보원 신뢰도 자체가 하락할 수 있다. 다만 공격자가 이미 강한 정체성 공동체 안에 있으면 외부의 폭로가 오히려 결속을 강화할 수 있으므로 내부적으로 신뢰받는 제 3 자·지역 언론·전문가의 역할이 중요하다.[13][14]

## 3.2 허위정보 폭로와 맥락 복원

사실검증은 날짜·장소·수치·원문·영상 출처·통계의 분모를 확인하는 기초 방어다. 그러나 “틀렸다”는 판정만 제시하지 말고 왜 틀렸는지, 무엇이 사실인지, 어떤 맥락이 빠졌는지를 함께 제공해야 정보의 빈 공간을 메울 수 있다.

## 3.3 프레임 해체

선전은 사실을 넘어서 “무엇이 문제인가, 누가 책임자인가, 어떤 감정을 느껴야 하는가, 무엇을 해야 하는가”를 한 묶음으로 제시한다. 따라서 대항선전은 상대의 문제정의와 인과관계를 분해하고, 누락된 변수와 대안적 해석을 보여주는 프레임 분석이 필요하다.

### 3.4 출처 검증

익명계정, 위장언론, 가짜 전문가, 조작된 인용은 메시지 내용만 검사해서는 잡히지 않는다. 원 출처, 최초 게시시각, 계정 생성패턴, 도메인 소유구조, 자금·조직 연계, 이미지·영상의 provenance 를 확인하는 것이 핵심이다. NIST 는 합성 콘텐츠 대응에서 provenance·labeling·detection·authentication 을 서로 보완적인 수단으로 제시한다.[22]

### 3.5 인지적 면역과 사전예방

개별 거짓말을 끝없이 따라다니는 방식은 공격자에게 의제를 맡긴다. 프리벙킹은 “감정적 언어, 가짜 전문가, 허위 양자택일, 음모적 출처, 맥락 잘라내기, 과도한 확신” 같은 조작기법을 사전에 학습시켜 새로운 거짓말에도 적용 가능한 일반 저항력을 만든다.[16][18]

### 3.6 대안적 서사 제공

사람은 기존 설명이 무너진 뒤에도 세계를 이해할 이야기를 필요로 한다. 반박만 남기면 인지적 공백이 생기므로, 사실에 부합하고 가치·정체성과 양립 가능한 대안적 설명을 제공해야 한다. 대테러 counter-narrative 연구에서도 “무엇에 반대하는지”만 말하지 말고 “무엇을 지지하는지”를 보여주는 positive/alternative narrative 가 강조되었다.[26][27]

### 3.7 사회적 신뢰와 제도적 정당성 보호

가장 장기적인 대항선전은 정부 홍보가 아니라 투명한 통계, 독립적 사법·감사, 자유로운 언론, 접근 가능한 공공데이터, 빠른 오류정정 같은 제도적 성능이다. 정부의 말과 행동 사이에 큰 차이가 생기면 어떤 메시지도 그 간극을 메우기 어렵다.[21][29]

## 4. 유형별 대항선전: 장점·한계·적용조건

| 유형 | 핵심 방식 | 장점 | 한계/역효과 | 적용조건 |
|---|---|---|---|---|
| 직접 반박 | 상대 주장을 명시적으로 인용·반박 | 빠르고 명료; 위기 대응에 적합 | 거짓 주장 재노출·의제 종속 | 확산 규모가 크고 피해가 즉각적일 때 |
| 간접 반박 | 상대 주장을 반복하지 않고 사실·기준을 제시 | 오정보 반복 최소화 | 대상이 무엇인지 청중이 이해 못할 수 있음 | 거짓말 자체가 아직 널리 알려지지 않았을 때 |
| 사실검증 | 증거·원문·데이터로 참/거짓 판단 | 검증 가능성·언론규범과 잘 맞음 | 정체성·감정·서사 효과는 남을 수 있음 | 검증 가능한 사실주장에 적합 |
| 출처 폭로 | 위장 출처·자금·조직·조작과정을 공개 | 신뢰도 기반 공격에 강함 | 증거가 약하면 명예훼손·역풍 | 명확한 attribution 증거가 있을 때 |
| 논리적 모순 제시 | 같은 정보원의 상충 주장·예측 실패를 비교 | 자기모순을 이용해 설득 부담 감소 | 정체성 높은 청중은 합리화 가능 | 기록이 충분하고 시간 비교가 가능할 때 |
| 리프레이밍 | 문제정의·인과·가치기준을 재구성 | 프레임 수준에서 경쟁 가능 | 과도한 정치화·메시지전으로 변질 위험 | 사실은 같지만 의미 해석이 경쟁할 때 |
| 대안 서사 | 긍정적·설명력 있는 대체 이야기 제공 | 인지적 공백을 메움; 장기적 | 서사가 사실보다 앞서면 또 다른 선전이 됨 | 정체성·가치 갈등이 큰 이슈 |
| 프리벙킹/접종 | 조작기법을 사전에 학습 | 확장성·예방효과 | 효과 지속·행동변화에 한계; 반복교육 필요 | 예측 가능한 조작기법·선거·재난·보건 위기 |
| 유머·풍자 | 과장·위선·권위 이미지를 희화화 | 공유성·기억성·권위 약화 | 모욕·오해·집단비하·핵심청중 미도달 | 문화코드가 공유되고 위험이 낮을 때 |
| 플랫폼 대응 | 라벨·노출감소·봇제거·광고투명성·계정조치 | 확산 인프라를 직접 건드림 | 검열·오탐·편향 논쟁; 투명성 필요 | 조직적 조작·가짜계정·불법콘텐츠가 확인될 때 |
| 미디어 리터러시 | 출처·증거·기법 판별 역량 교육 | 장기 회복탄력성 | 효과가 느리고 교육격차 존재 | 상시적 시민교육·학교·언론교육 |

### 4.1 “응답하지 않기”도 전략이다

모든 선전에 반응할 필요는 없다. 작은 허위주장에 정부·대형언론이 직접 대응하면 오히려 도달률과 정당성을 부여할 수 있다. 대응 여부는 최소한 ① 실제 도달 규모, ② 취약 청중에 대한 침투, ③ 피해의 심각성, ④ 확산속도, ⑤ 교정 가

능한 증거의 존재, ⑥ 대응 자체의 증폭효과를 평가한 뒤 결정해야 한다. “무시-관찰-간접교정-직접반박-플랫폼조치”를 연속선으로 보고 강도를 조절하는 것이 바람직하다.

## 5. 심리학적 메커니즘: 왜 사실만으로는 부족한가

대항선전은 정보정확성의 문제가 아니라 인간의 기억·주의·정체성·사회적 관계의 문제이기도 하다. 사람들은 모든 정보를 독립적으로 재검증하지 않고, 반복성·친숙성·출처·집단규범·감정강도를 휴리스틱으로 사용한다. 따라서 사실 검증 자체는 필요조건이지만 충분조건이 아니다.

| 메커니즘 | 대항선전에 미치는 영향 | 대응 원리 |
|---|---|---|
| Familiarity effect | 반복해서 본 문장은 처리하기 쉬워져 친숙하고 그럴듯하게 느껴질 수 있다. 거짓말을 반박하면서 제목과 이미지를 반복하면 의도치 않게 기억 흔적을 강화할 수 있다. | “거짓을 길게 반복→마지막에 정정”보다 사실을 먼저 제시하고 거짓주장은 필요한 만큼만 짧게 언급한다. |
| Backfire effect | 교정이 오히려 오정보 믿음을 높이는 현상. 중요한 위험이지만 최근 연구에서는 보편적·상시적 현상으로 재현되지 않으며 특정 항목·청중·측정조건에서 나타나는 것으로 보는 시각이 강하다.[17] | 역효과 가능성을 이유로 교정을 포기하지 말고, 신뢰받는 출처·비위협적 언어·대안설명·반복 측정으로 관리한다. |
| Motivated reasoning | 사람은 원하는 결론을 지지하는 증거에는 관대하고 위협하는 증거에는 엄격할 수 있다. | 정체성 공격을 피하고 공유가치·절차·증거기준을 강조한다. |
| Confirmation bias | 기존 믿음과 일치하는 정보에 더 주목·수용하고 반대 정보는 무시하는 경향. | 청중이 이미 신뢰하는 채널과 제 3 자 정보원을 활용하고 다양한 증거경로를 제공한다. |
| Identity-protective cognition | 사실판단이 집단정체성과 결합하면 “사실 인정=배신”처럼 느껴질 수 있다. | 정체성을 굴복시키는 방식이 아니라 존엄·소속을 유지하면서 사실을 수정할 수 있는 출구를 만든다. |
| Selective exposure | 사람은 선호하는 정보환경을 선택하고 불편한 정보는 회피한다. | 상대 진영 공간에 억지로 침투하기보다 교차노출이 가능한 중간자·지역언론·생활정보 채널을 활용한다. |
| Echo chamber | 동질적 네트워크가 반복노출과 사회적 증거를 강화한다. | 콘텐츠 교정과 함께 네트워크 다양성·추천시스템·가짜계정 문제를 함께 본다. |
| Group polarization | 동질적 집단 토론이 평균 의견을 더 극단화할 수 있다. | 공개 논쟁으로 상대 지지자를 집결시키기보다 비공개·숙의형·교차집단 접촉을 고려한다. |
| Source credibility | 전문성·정직성·독립성·청중과의 관계가 메시지 수용을 결정한다.[6] | “정부가 말했으니 믿어라”보다 검증 가능한 증거와 독립기관의 확인을 우선한다. |
| Social proof | 많은 사람이 믿고 공유하는 것처럼 보이면 신뢰가 올라갈 수 있다. 봇·조정계정은 이 신호를 인위적으로 조작한다. | 좋아요·리트윗 수보다 계정 품질·확산 구조·조정행동을 분석한다. |

**연구상의 주의** 심리효과의 크기는 이슈·문화·정치정체성·측정시점에 따라 달라진다. 단일 실험의 효과크기를 모든 국가·플랫폼에 일반화해서는 안 된다. 특히 태도변화와 실제 행동변화는 구분해야 한다.

## 6. 역사적 발전

**대항선전의 역사적 발전: 매체 중심 대응에서 정보생태계 회복탄력성으로**

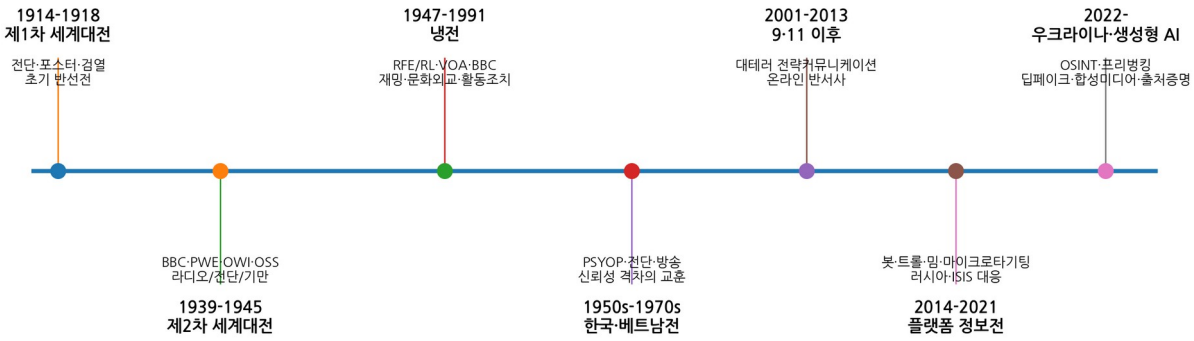

그림 2. 대항선전의 역사적 발전 개관

## 6.1 제 1 차 세계대전: 대중선전의 산업화와 초기 반선전

총력전은 신문, 포스터, 영화, 전단, 우편, 강연을 국가전략의 일부로 만들었다. 영국의 Wellington House, 미국의 Committee on Public Information, 독일의 전시 선전기관 등은 국내 동원과 적국·중립국 설득을 병행했다. 이 시기의 "대항"은 오늘날처럼 독립된 전문영역이라기보다 적의 atrocity story·전황 주장·전쟁목표를 반박하고 검열·언론통제로 내부 사기를 유지하는 방식이었다. 전쟁 후 과장·조작 논란이 드러나면서 "선전"이라는 말 자체가 부정적 의미를 강하게 갖게 되었고, 이는 제 2 차 세계대전에서 연합국이 "정보"와 "진실"을 강조하는 배경이 되었다.[3]

## 6.2 제 2 차 세계대전: 심리전의 전문화

영국 Political Warfare Executive(PWE), BBC, 미국 OWI 와 OSS, 독일 RMVP, 소련의 선전기관은 라디오·전단·소문·위장매체를 대규모로 운용했다. 영국의 Sefton Delmer 가 운영한 비밀 라디오들은 내부 불만과 나치 엘리트의 위선·부패를 활용하여 독일 청중을 겨냥했다. 이는 청중의 기존 가치와 언어를 이해한 정교한 counter-propaganda 였지만, 허위·위장 출처를 사용한 "black propaganda"라는 윤리적 문제도 남긴다.[36] 소련도 독일군을 대상으로 한 전단과 Front-Illustrierte 같은 매체로 나치 선전을 공격했다. 이 시기 핵심 교훈은 "상대의 메시지에 답하는 것"보다 상대 사회의 균열과 신뢰구조를 이해하는 것이 중요하다는 점이다.

## 6.3 냉전: 장기 방송·문화외교·활동조치의 경쟁

냉전기에는 RFE/RL, VOA, BBC World Service 와 소련 Radio Moscow, 국제방송, 문화교류, 출판, 학술·예술 후원, 전단, 비밀 정치활동이 결합했다. 서방 방송은 폐쇄된 미디어 환경에 대안 뉴스를 제공했고, 소련 및 동유럽 국가들은 재밍·검열·형사처벌로 대응했다. 동시에 소련의 "active measures"는 위조문서·전위조직·유언비어·언론플레이를 활용했다. 이 경쟁은 대항선전이 단기 반박이 아니라 장기간 신뢰를 축적하는 "대체 정보 인프라"가 될 수 있음을 보여준다.

## 6.4 한국전쟁과 베트남전쟁: 작전효과와 신뢰성 격차

한국전쟁에서 미군·유엔군과 북한·중국은 전단, 확성기, 포로 관련 메시지, 항복 권고, 전황선전을 광범위하게 사용했다. 베트남전쟁에서는 미국의 JUSPAO, Chieu Hoi 프로그램, 라디오·전단과 북베트남·베트콩의 선전이 경쟁했다. 그러나 전황·정책에 대한 공식 메시지와 현장 경험이 어긋나면서 이른바 "credibility gap"이 커졌고, 대항선전의 신뢰는 커뮤니케이션 기술보다 실제 정책과 성과에 의존한다는 교훈을 남겼다.

## 6.5 탈냉전과 9·11 이후: 전략커뮤니케이션과 대테러 counter-narrative

위성방송과 인터넷이 등장하면서 중앙집중형 방송전은 다중 플랫폼 경쟁으로 바뀌었다. 9·11 이후 알카에다와 ISIS 가 동영상·포럼·SNS 를 활용하자 미국·유럽·중동 정부는 대테러 전략커뮤니케이션과 counter-narrative 를 발전시켰다. 그러나 정부가 직접 극단주의자와 논쟁하는 방식은 정보원 신뢰도와 문화적 거리 때문에 한계를 보였다.
ICCT/Hedayah 연구는 지역의 신뢰받는 목소리, positive/alternative narrative, 청중 세분화, 온라인·오프라인 결합을 강조했다.[26][27]

## 6.6 2010 년대 러시아 정보전과 플랫폼 시대

러시아의 정보활동에 관한 연구는 다수 채널, 높은 메시지량, 빠른 반복, 사실성에 대한 낮은 헌신, 서로 모순되는 주장까지 병렬적으로 활용하는 "firehose of falsehood" 특성을 지적했다.[11] SNS 에서는 국가매체뿐 아니라 트롤·봇·가짜계정·해시태그 캠페인·지역적 이슈 증폭이 결합했다.[12] 이에 대한 대응은 단순 반박보다 신뢰받는 지역언론, 독립 콘텐츠 생산자, 미디어 리터러시, 조정행동 탐지, 플랫폼 투명성을 강조하는 방향으로 이동했다.[13][14]

## 6.7 ISIS 온라인 선전과 대응

ISIS 는 "칼리프 국가의 성공·소속감·영웅성·폭력의 충격성"을 고품질 영상·잡지·밈·개인 네트워크와 결합했다. 미국의 Think Again Turn Away 는 ISIS 의 잔혹성과 위선을 직접 반박했지만, 정부라는 발신자의 낮은 신뢰도, 공개논쟁이

상대 콘텐츠를 증폭할 위험, 성과 측정의 어려움으로 비판받았다.[34][35] 이후 접근은 제 3 자 목소리, 탈퇴자·피해자 증언, 검색 리디렉션, 커뮤니티 참여, 대안 서사로 이동했다.

### 6.8 우크라이나 전쟁과 오픈소스 검증

2022 년 이후 전쟁은 정부 발표, 위성영상, 시민영상, OSINT, 텔레그램, 틱톡, X, 유튜브가 동시에 경쟁하는 실시간 정보전이 되었다. 사전 정보공개와 false-flag 가능성에 대한 경고는 "프리벙킹"의 국가안보 적용 사례로 논의되었으나, 전시 정부의 모든 주장을 자동으로 사실로 취급해서는 안 된다. 독립적 영상검증, 지리위치·시간확인, 위성자료, 여러 출처의 교차확인이 중요하다.

### 6.9 생성형 AI 시대

생성형 AI 는 텍스트·이미지·음성·영상 제작비용을 낮추고, 다국어·다인격·고빈도 콘텐츠를 자동화할 수 있다. CISA 는 2024 선거 관련 평가에서 생성형 AI 가 완전히 새로운 위험을 만든다기보다 기존 위험을 증폭할 가능성이 크다고 보았고,[23] NIST 는 provenance, watermarking/labeling, detection, authentication 이 모두 불완전하므로 사회·제도적 대책과 결합해야 한다고 강조한다.[22][37]

## 7. 국가·체제별 대항선전·정보대응 체계 비교

| 국가/체제 | 주요 체계·역사 | 특징 | 비판적 평가 |
|---|---|---|---|
| 미국 | OWI/OSS, USIA, VOA/RFE-RL, DoD PSYOP/MISO, 공공외교, GEC(2016-2024) 등 | 공개·비공개 수단 혼재; 정부와 편집독립 방송 사이의 제도적 긴장 | 정보원 신뢰·다원성 장점 vs 정권 변화·정치화·국내 표현자유 논쟁. GEC 는 2024 년 종료, 후속 조직도 2025 년 폐쇄되어 제도 변동성이 큼.[28][29] |
| 영국 | BBC External/World Service, PWE, 정부 전략커뮤니케이션 | 장기 국제방송 신뢰와 전시 black propaganda 의 이중 유산 | BBC 의 편집규범은 강점; 전시 기만은 민주적 투명성과 긴장 |
| 소련/러시아 | 당-국가 선전, 검열·재밍, active measures, 국제방송, 현대 국가매체·온라인 네트워크 | 내부 정보통제와 외부 영향작전의 결합 | 상대의 모순 증폭·다중서사·불신 조성이 강점; 사실성 부족은 장기 신뢰 취약점.[11][12] |
| 중국 | 당 선전·외선(外宣), 국영 국제매체, 플랫폼 관리, 외교·문화·디지털 영향 | 당-국가와 미디어·플랫폼의 높은 통합 | 자원·조정력 강점; 출처 독립성과 검열 때문에 해외 신뢰 문제. 연구는 중국의 국가 메시지와 covert influence 를 구분해야 함. |
| 나치 독일 | Reich Ministry of Public Enlightenment and Propaganda, 검열, 라디오·영화·신문 통제, 적 방송 금지 | 전체주의적 정보독점과 적 선전 억제 | 대항선전이 사실검증이 아니라 정권 서사 유지와 강제 통제에 종속됨 |
| 북한 | 당 선전선동체계, 외부정보 차단·재밍, 대남·대외 선전, 내부 비판서사 | 극도로 폐쇄된 정보환경에서 "대항"이 검열·이념통제와 결합 | 독립검증·언론자유·출처투명성 부재로 민주적 counter-propaganda 와 본질적으로 다름 |

비교에서 중요한 것은 "민주국가는 선전하지 않고 권위주의국가는 선전한다"는 도덕적 이분법을 피하는 것이다. 민주국가도 전쟁과 대외정책에서 설득·심리전·기만을 사용해 왔고, 권위주의국가의 모든 정보가 거짓인 것도 아니다. 분석자는 개별 주장마다 증거를 검증하고, 제도적으로는 출처 공개·독립 언론·사법심사·감사·정정 가능성 등 견제장치가 존재하는지를 평가해야 한다.

## 8. 주요 학자와 이론이 대항선전에 주는 통찰

| 학자 | 대항선전 이해에 대한 기여 |
|---|---|
| Harold Lasswell | 선전을 상징을 통한 집단태도 관리와 전쟁 동원의 기술로 분석. "누가 무엇을 어떤 채널로 누구에게 어떤 효과로 말하는가"라는 커뮤니케이션 분석틀의 선구적 토대.[3] |
| Jacques Ellul | 현대 선전은 개별 거짓말보다 사회의 기술·조직·대중매체 환경에 의존한다고 분석. 대항선전도 메시지 하나가 아니라 전체 사회환경과 개인의 통합 욕구를 봐야 함.[2] |

| | |
|---|---|
| Edward Bernays | PR 과 여론형성의 조직적 기법을 체계화. 대항선전 연구에는 의제설정·권위있는 제 3 자·상징 사용의 효과와 윤리적 위험을 동시에 보여줌.[4] |
| Walter Lippmann | 사람이 직접 세계를 경험하기보다 "머릿속의 그림"을 통해 판단한다는 관점. 프레임·스테레오타입·매개된 현실의 중요성을 설명.[5] |
| Paul Lazarsfeld | 2 단계 흐름(two-step flow)과 opinion leader 연구는 대중매체의 직접효과보다 사회적 관계와 중간자의 중요성을 강조. 오늘날 인플루언서·커뮤니티 허브 분석으로 연결. |
| Carl Hovland | 메시지·정보원·청중 특성에 따른 설득 연구. source credibility, sleeper effect 등은 왜 신뢰받는 발신자가 중요한지 설명.[6] |
| Leon Festinger | 인지부조화 이론은 기존 정체성과 충돌하는 정보가 거부·합리화될 수 있음을 설명.[7] |
| William McGuire | inoculation theory 를 발전시켜 약한 공격과 반박을 미리 경험하면 이후 설득에 대한 저항력이 커질 수 있음을 제시.[8] |
| Daniel Kahneman & Amos Tversky | 휴리스틱·편향·프레이밍·위험지각 연구는 정보판단이 순수한 논리계산이 아니라는 점을 보여준다. 대항선전은 단순 정보량 증가가 아니라 판단환경 설계를 고려해야 함.[9] |

# 9. 대항선전의 9 개 분석 수준

그림 3. 메시지에서 제도까지: 대항선전의 분석 단위

| 수준 | 효과를 좌우하는 변수 | 분석상 주의 |
|---|---|---|
| 메시지(message) | 주장 정확성, 증거의 질, 명료성, 반복, 감정강도, 이미지·상징, 반론 포함 여부 | A/B 테스트는 가능하지만 단기 클릭률을 설득효과로 오해하지 말 것 |
| 정보원(source) | 전문성, 정직성, 독립성, 과거 정확도, 정체성 적합성, 이해상충 | "누가 말했는가"가 사실내용만큼 중요할 수 있음 |
| 매체(channel) | 도달률, 형식 적합성, 속도, 편집규범, 접근성, 언어·문화 적합성 | 라디오·TV·메신저·쇼츠는 신뢰구조와 사용맥락이 다름 |
| 청중(audience) | 사전신념, 정치·종교·집단정체성, 지식, 감정, 미디어 습관, 취약성 | 청중을 조작대상이 아니라 자율적 판단주체로 취급해야 민주적 정당성이 유지됨 |
| 프레임(frame) | 문제 정의, 인과귀속, 도덕평가, 해결책, 책임주체 | 팩트는 맞아도 상대의 프레임 안에서 답하면 의제에 종속될 수 있음 |
| 서사(narrative) | 주인공·악당·피해자, 역사적 기억, 시간구조, 공동체의 의미 | 대안서사는 사실성·복잡성·정정 가능성을 유지해야 함 |
| 네트워크(network) | 허브, 브리지, 커뮤니티 구조, coordinated behavior, social proof | 콘텐츠 삭제보다 조정 네트워크 해체가 더 효과적일 수 있음 |
| 알고리즘(algorithm) | 추천·랭킹·광고·트렌드·자동완성·검색 노출 | 설계가 반복노출과 감정적 콘텐츠의 보상을 바꿀 수 있음 |
| 제도(institution) | 법, 감사, 선거관리, 언론자유, 공공데이터, 교육, 사법·의회 통제, 신뢰 | 장기 회복탄력성의 최종 기반 |

# 10. 디지털 플랫폼·봇·딥페이크·생성형 AI 가 바꾸는 구조

## 10.1 봇과 트롤: 가짜 사회적 증거

봇은 메시지를 자동 게시·재게시하고, 트롤·조정계정은 댓글·논쟁·해시태그를 통해 특정 견해가 "많은 사람이 믿는 것"처럼 보이게 만들 수 있다. 핵심 효과는 개별 메시지의 설득보다 visibility, repetition, agenda salience, social proof 조작이다. 따라서 계정 단위 판별만으로는 부족하고 시간동기화, 콘텐츠 유사성, 공유 URL, 계정군의 상호작용 패턴 등 coordinated behavior 를 봐야 한다.[12][14]

### 10.2 딥페이크와 합성미디어: “거짓의 증거”와 “진짜의 부정”

합성영상·음성은 존재하지 않은 사건을 만들어낼 뿐 아니라, 실제 증거가 공개되었을 때 “AI 로 만든 것”이라고 부인할 수 있는 liar’s dividend 를 키울 수 있다. 방어는 탐지기 하나에 의존하기보다 촬영·편집 이력, cryptographic provenance, 원본 보관, 독립 출처, 현장 증거를 결합해야 한다. NIST 도 단일 기술이 완전한 해결책이 아니라고 강조한다.[22]

### 10.3 알고리즘 추천과 플랫폼 증폭

추천시스템은 정치적 의도를 갖지 않더라도 참여·시청시간 최적화 과정에서 감정적·논쟁적 콘텐츠를 보상할 수 있다. 대항선전은 “좋은 콘텐츠 생산”만이 아니라 추천 시스템의 위험평가, 투명성, 연구자 접근, 광고라이브러리, 계정조정 규칙과 연결된다. EU DSA 는 대형 플랫폼의 선거·시민담론 관련 시스템 위험을 식별·완화하도록 요구하는 방향을 제도화했다.[31]

### 10.4 밈과 인플루언서

밈은 압축된 감정·정체성·내집단 신호를 빠르게 전파한다. 인플루언서는 공식기관보다 높은 관계적 신뢰를 가질 수 있으나, 정부가 사실상 대리발신자를 은폐하면 astroturfing·선전 논란이 발생한다. 지원관계·광고·협업의 투명성이 핵심이다.

### 10.5 마이크로타기팅

개인 데이터와 행동추정에 기반한 미세타기팅은 동일한 정치행위자가 서로 다른 집단에 상충하는 메시지를 보여주며 공적 검증을 회피할 수 있게 한다. 민주적 대응은 타기팅 자체의 “효율”보다 광고 투명성, 민감정보 사용 제한, 공개 아카이브, 감사가능성을 중시해야 한다.

### 10.6 온라인 커뮤니티

폐쇄형 메신저·포럼·팬덤·이념 공동체에서는 정보가 사회적 관계와 정체성 규범을 통해 검증된다. 외부 팩트체커가 정면개입하면 “외집단 공격”으로 해석될 수 있으므로 내부의 신뢰받는 중간자와 공동체 규범이 중요하다.

### 10.7 생성형 AI 와 에이전트형 영향작전

생성형 AI 는 문체·언어·인격을 빠르게 변형하고 댓글 응답까지 자동화할 수 있어 “한 캠페인-다수 페르소나” 운영비용을 크게 낮춘다. 동시에 AI 는 번역·팩트체크 보조·콘텐츠 출처 분석·네트워크 이상탐지에도 활용될 수 있다. 방어의 초점은 모델 자체 차단보다 campaign coordination, provenance, rate limits, identity abuse, audit trails, 사회적 신뢰로 이동할 가능성이 크다.[37][38]

## 11. 민주주의·법·윤리·인권: 대항선전과 검열의 경계

그림 4. 민주적 대항선전과 권위주의적 정보통제를 구분하는 규범적 경계

민주주의 국가의 대항선전은 역설을 갖는다. 시민의 자유로운 판단을 보호하기 위해 허위정보와 영향작전에 대응하지만, 정부가 “허위”의 정의와 정보 유통경로를 독점하면 그 대응 자체가 검열·정부선전·정치적 반대 억압으로 변할 수 있다. 따라서 민주적 정당성은 목적보다 절차와 견제에서 나온다.

| 원칙 | 민주적 기준 |
|---|---|
| 투명성 | 정부·군·기관이 발신자인지 공개. 제 3 자·인플루언서 지원도 이해상충을 표시. |

| | |
|---|---|
| 사실성 | 확인된 사실과 추정·평가·정보기관 판단을 명확히 구분. 오류 발견 시 신속 정정. |
| 출처 공개 | 가능한 범위에서 원자료·방법론·근거를 제공하고 보안상 비공개 근거는 그 한계를 명시. |
| 독립적 검증 | 언론·학계·사법·감사기관이 정부 주장과 플랫폼 조치를 독립적으로 검증할 수 있어야 함. |
| 비례성 | 피해 규모에 맞는 최소한의 제한수단. 반박·라벨·노출조정·계정제재·삭제·형사처벌은 동일한 단계가 아님. |
| 법적 통제 | 명확한 법률근거, 사법심사, 이의제기 절차, 기록보존, 의회·감사 통제를 확보. |
| 언론 자유 | 정부에 불리하거나 불편한 보도, 오류 가능성이 있는 주장까지 폭넓게 보호. 비판언론을 "적 선전"으로 낙인찍지 않음. |
| 시민의 자율성 | 국민을 조작의 객체가 아니라 정보를 비교·판단할 권리를 가진 주체로 대우. |
| 프라이버시 | 대항선전 명목의 감시·프로파일링·마이크로타기팅은 필요성·비례성·법적 절차를 충족해야 함. |

### 11.1 국제법·인권법

ICCPR 제 19 조는 의견 보유와 표현의 자유를 보호하며, 표현 제한은 법률에 의해 정해지고 타인의 권리·명예, 국가안보, 공공질서, 공중보건·도덕 등 정당한 목적을 위해 필요하고 비례적이어야 한다.[30] 제 20 조는 전쟁선전과 차별·적대·폭력을 선동하는 특정 유형의 혐오선동을 법률로 금지하도록 요구한다. 따라서 "허위정보"라는 이유만으로 광범위한 형사처벌·차단을 정당화할 수 없으며, 내용의 위험성·의도·맥락·도달·현실적 위해 가능성을 정교하게 평가해야 한다. UN 인권기준은 과도하게 넓은 허위정보법이 정치적 비판과 언론을 위축시킬 수 있음을 반복해서 경고해 왔다.[30]

### 11.2 국내법과 플랫폼 규칙

국내법은 국가마다 다르다. 미국처럼 정부의 표현규제를 강하게 제한하는 체계와, 선거 허위정보·혐오표현·플랫폼 시스템 위험에 더 적극적으로 규율하는 유럽 체계는 접근법이 다르다. 플랫폼의 커뮤니티 규칙은 헌법과 동일하지 않지만, 대형 플랫폼이 사실상 공론장의 인프라가 된 만큼 규칙의 명확성·일관성·이의제기·투명성·연구자 접근성이 중요하다. EU DSA 는 대형 플랫폼의 시스템 위험을 규제하면서도 기본권 보호를 결합하려는 하나의 모델이다.[31]

### 11.3 언론윤리

언론의 대항선전은 정부 메시지를 대신 전하는 것이 아니라 검증하는 데 있다. 핵심 원칙은 원문·원자료 확인, 헤드라인에서 거짓주장의 과도한 반복 회피, 출처와 이해상충 표시, 정정의 가시성, 불확실성 공개, 조작영상의 불필요한 재유통 최소화, 독립적 설명과 맥락 제공이다. UNESCO 의 저널리즘·허위정보 교육 자료는 fact-checking, social media verification, media and information literacy 를 이러한 전문규범과 연결한다.[20]

## 12. 효과 평가: 조회수는 설득효과가 아니다

| 지표 | 측정 대상 | 주의점 |
|---|---|---|
| 도달률 Reach | 목표청중 중 실제 노출된 비율 | 노출만으로 태도·행동변화를 뜻하지 않음 |
| 주의·완독 Attention | 시청시간, 완독률, 재방문 | 높은 관심이 지지인지 분노·조롱인지 구분 필요 |
| 신뢰도 변화 Credibility | 정보원·기관에 대한 신뢰 전후 변화 | 단기 메시지 성과가 장기 신뢰손실을 가릴 수 있음 |
| 사실신념 변화 Belief accuracy | 구체 주장에 대한 정확도 | 정답암기와 전이효과를 구분 |
| 태도 변화 Attitude | 정책·집단·기관에 대한 태도 | 사회적 바람직성 편향·설문효과 주의 |
| 행동 변화 Behavior | 공유, 신고, 투표정보 확인, 가입·이탈 등 | 인과추론이 어려우며 윤리적 개인정보 보호 필요 |
| 허위정보 확산률 | 재전파율, reproduction number 유사 지표, 노출속도 | 플랫폼 데이터 접근과 정의가 필요 |
| 네트워크 전파 | 허브·브리지·커뮤니티간 이동, 조정계정 변화 | 콘텐츠 수준 지표보다 캠페인 구조 파악에 유용 |
| 정보원 신뢰도 | 장기 정확도·정정 기록·독립성 | 단기 viral success 보다 전략적으로 중요 |

| | | |
|---|---|---|
| 회복탄력성 Resilience | 새로운 오정보에 대한 판별력, 신뢰 회복, 기관 대응속도 | 장기 패널·반복 측정 필요 |

가장 흔한 오류는 “조회수 100 만 = 100 만 명 설득”으로 해석하는 것이다. 조회는 호기심·반대·조롱·봇·중복노출을 포함한다. 효과 평가는 최소한 노출 → 이해 → 신뢰 → 사실신념 → 태도 → 행동의 단계를 구분하고, 가능하면 사전-사후 설계, 대조군, 장기추적, 네트워크 분석을 결합해야 한다. RAND 의 공공커뮤니케이션 평가 연구도 메시지의 성과를 하나의 engagement 지표로 환원하지 말 것을 강조한다.[15]

# 13. 성공·실패 사례 비교

## 13.1 상대적 성공 사례: 폐쇄사회에 대한 장기 국제방송

BBC World Service, VOA, RFE/RL 같은 국제방송은 냉전기 폐쇄된 정보환경에서 대안 뉴스와 문화·사회 정보를 지속적으로 제공했다. 이를 “공산권 붕괴의 원인”으로 단순화할 수는 없지만, 장기간의 정기성, 현지어 전문성, 청취자 요구에 맞는 프로그램, 대체 정보원으로서의 신뢰 축적은 대항선전의 구조적 성공요인을 보여준다. 성공의 핵심은 하루의 반박 캠페인이 아니라 “상대가 숨기는 정보가 있을 때 확인할 수 있는 습관적 채널”을 만드는 것이었다. 반대로 이 방송들이 국가정책과 연결되어 있었다는 사실은 편집독립성과 선전의 경계에 대한 지속적 논쟁을 낳았다.

| 요소 | 평가 |
|---|---|
| 목표 | 폐쇄된 미디어 환경에 지속적 대안 정보 공급 |
| 청중 | 동유럽·소련 등 현지 청취자 |
| 메시지 | 뉴스·해설·문화·생활정보의 혼합 |
| 정보원 | 국가재정 지원이지만 방송별 편집규범·전문성 구축 |
| 전달매체 | 단파·중파 라디오, 이후 디지털 |
| 타이밍 | 위기 때만이 아니라 일상적·장기적 |
| 성과 | 대체 정보접근과 신뢰의 축적; 정치변동의 단독 원인으로 과장해서는 안 됨 |
| 핵심교훈 | 신뢰는 반복되는 정확성과 생활 속 유용성에서 형성 |

## 13.2 실패/경고 사례 1: RFE 와 헝가리 혁명 1956

1956 년 헝가리 혁명에서 RFE 방송의 역할은 오랫동안 논쟁적이다. RFE/RL 자체의 후속 연구는 방송이 혁명을 “일으켰다”거나 서방 군사개입을 명시적으로 약속했다는 강한 주장을 부정한다.[32] 반면 미국 National Security Archive 가 소개한 자료와 일부 연구는 혁명 중 일부 방송의 공격적 어조와 전술적 조언이 청취자에게 서방 지원에 대한 과도한 기대를 줄 수 있었다고 비판한다.[33] 가장 합리적인 평가는 “직접 선동의 단순 인과론”을 피하면서도, 위기상황에서 모호한 기대를 만들 수 있는 메시지 책임을 인정하는 것이다.

- 실패요인: 위기 시 청중의 희망적 해석, 불명확한 정책선, 방송 어조와 실제 서방 행동의 불일치.
- 교훈: 대항선전은 청중이 무엇을 “추론할지”까지 고려해야 하며, 전달자가 실제로 제공할 수 없는 지원을 암시해서는 안 된다.

## 13.3 실패/경고 사례 2: Think Again, Turn Away

미 국무부의 Think Again, Turn Away 는 2013 년 이후 알카에다·ISIS 선전에 직접 대응한 소셜미디어 캠페인이었다. 잔혹행위·위선·전장실패를 보여주는 콘텐츠는 상당한 조회를 얻기도 했지만, 핵심 청중에게 미국 정부가 신뢰받는 발신자인지, 공개적인 “트롤링”이 오히려 ISIS 메시지를 증폭하는지, 조회수가 실제 이탈·비가입으로 이어지는지에 대한 비판이 컸다.[34][35] 이후 counter-narrative 분야에서는 정부의 직접 논쟁보다 지역·민간의 신뢰받는 목소리, 대안서사, 커뮤니티 참여, 검색·추천단계의 개입이 강조되었다.[26][27]

| 비교요소 | 장기 국제방송 모델 | Think Again, Turn Away |
|---|---|---|
| 핵심 논리 | 지속적 대체 정보환경 | 상대 메시지에 직접 반박·공격 |

| | | |
|---|---|---|
| 정보원 | 전문방송·편집규범 | 미국 정부가 명시적 발신자 |
| 청중 관계 | 일상적 청취 습관과 장기 관계 | 적대적·위험 청중과 공개 논쟁 |
| 성과측정 | 장기 신뢰·접근성 | 조회수·공유 등 단기 지표가 눈에 띄기 쉬움 |
| 주요 위험 | 국가재정과 편집독립성 논쟁 | 발신자 불신·상대 증폭·효과측정 불확실성 |

# 14. 통합 모델과 실무적 진단 프레임

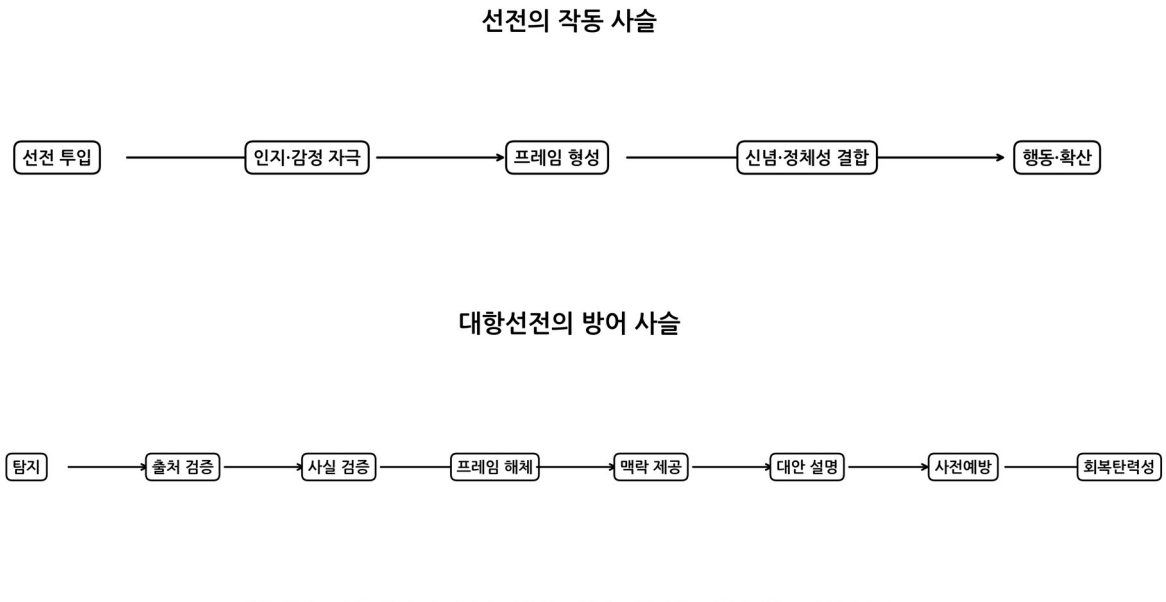


그림 5. 통합 모델: 선전의 작동 과정과 대항선전의 단계별 방어

통합모델의 핵심은 각 단계에서 서로 다른 방어수단이 필요하다는 것이다. "선전 투입" 단계에서는 출처·네트워크 탐지가, "인지·감정 자극" 단계에서는 감정적 조작기법에 대한 프리벙킹이, "프레임 형성" 단계에서는 누락된 맥락과 대안 프레임이, "신념·정체성 결합" 단계에서는 신뢰받는 내부자와 존엄을 보존하는 교정이, "행동·확산" 단계에서는 플랫폼·네트워크 차원의 마찰(friction)과 조정행동 차단이 중요하다.

| 단계 | 핵심 질문 |
|---|---|
| 1. 탐지 | 무엇이 급속히 확산되는가? 자연발생인가 조정행동인가? |
| 2. 출처검증 | 최초 출처·자금·계정·도메인·미디어 provenance 는 무엇인가? |
| 3. 사실검증 | 검증 가능한 주장은 무엇이며 증거 수준은 어느 정도인가? |
| 4. 프레임 해체 | 어떤 문제정의·원인·도덕평가·해결책을 전제하는가? |
| 5. 맥락 제공 | 무엇이 빠졌으며 시간·비교집단·분모·역사적 맥락은 무엇인가? |
| 6. 대안적 설명 | 거짓을 제거한 뒤 청중이 사용할 설명모델은 무엇인가? |
| 7. 사전예방 | 같은 조작기법을 다음에도 알아볼 수 있도록 일반화 가능한 교훈을 주는가? |
| 8. 회복탄력성 | 독립언론·교육·공공데이터·플랫폼 투명성·법적 통제를 강화하는가? |

## 14.1 대응 강도 결정 매트릭스

| 상황 | 권장 기본 대응 |
|---|---|
| 도달 낮음 + 위해 낮음 | 관찰, 검색결과에 사실정보 준비, 불필요한 직접 반박 회피 |

| | |
|---|---|
| 도달 높음 + 사실검증 용이 | 사실 우선 정정 + 간결한 거짓언급 + 원자료 링크 + 신뢰받는 제 3 자 |
| 조정된 봇/가짜계정 캠페인 | 콘텐츠 논쟁보다 네트워크·플랫폼 무결성 조치와 attribution |
| 정체성 결합이 강한 서사 | 직접 공격보다 내부 신뢰자·대안서사·숙의·커뮤니티 기반 개입 |
| 딥페이크/합성미디어 | 원본·provenance·현장증거·복수독립출처 확인, 탐지기 결과는 보조자료로 사용 |
| 정부 자체 신뢰위기 | 홍보 확대보다 데이터 공개·감사·정정·정책행동의 일치가 우선 |

## 15. 결론: 가장 강한 대항선전은 신뢰 가능한 정보질서다

대항선전의 역사는 전단과 라디오에서 소셜미디어, 딥페이크, 생성형 AI 로 매체가 바뀌어 온 역사이지만, 성공과 실패를 가르는 원리는 놀랄 만큼 일관된다. 사람은 “사실”만 믿는 것이 아니라 정보원, 사회적 관계, 정체성, 반복, 프레임, 제도적 경험을 함께 평가한다. 따라서 대항선전은 상대의 거짓말보다 더 세련된 조작을 만드는 경쟁이 되어서는 안 된다.

효과적인 대항선전의 1 차 기반은 정확한 정보다. 그러나 정확성만으로는 충분하지 않다. 그 정보가 신뢰할 수 있는 방식으로 생산되고, 출처가 투명하며, 독립적 검증이 가능하고, 오류가 공개적으로 정정되며, 시민이 다른 정보원을 비교할 자유가 있어야 한다. 독립언론, 공공데이터, 사법·의회·감사기관, 학계, 시민사회, 플랫폼 투명성, 미디어 리터러시는 각각 별개의 보조수단이 아니라 같은 방어체계의 구성요소다.[20][21][30]

궁극적으로 정보공간의 경쟁력은 “얼마나 빨리 반박했는가”가 아니라 “거짓이 얼마나 오래 살아남기 어려운가, 시민이 스스로 검증할 수 있는가, 제도가 오류를 인정하고 고칠 수 있는가, 위기 이후에도 신뢰가 회복되는가”로 측정해야 한다. 이 관점에서 대항선전은 선전의 반대편에 놓인 또 하나의 선전이라기보다, 민주사회가 사실성·검증 가능성·자율성을 지키기 위해 구축하는 장기적인 정보 회복탄력성의 체계라고 정리할 수 있다.

# 부록 A. 개념 비교표: 기능적 경계

| 개념 | 방향 | 핵심수단 | 성공조건 | 대표 위험 |
|---|---|---|---|---|
| 대항선전 | 방어/상쇄 | 반박·출처·프레임·서사·네트워크·교육 | 신뢰·증거·청중 적합성 | 역증폭·정부선전화 |
| 공공외교 | 관계/장기 | 대화·교류·문화·정책설명 | 상호성·지속성 | 정책-메시지 불일치 |
| 전략커뮤니케이션 | 조직통합 | 행동·정책·메시지 정렬 | 조직 일관성 | 홍보가 전략을 대체 |
| MISO/PSYOP | 작전지원 | 표적청중 분석·메시지·행동 | 정보·문화 이해 | 과도한 조작성 |
| 정보작전 | 작전통합 | 사이버·전자전·기만·정보활동 조정 | 전영역 통합 | 개념 과잉확장 |
| 인지전 | 인지공간 경쟁 | 주의·판단·정체성·사회인지 | 다학제 분석 | 정의 모호·군사화 |
| 팩트체크 | 교정 | 증거 검증 | 원자료·방법론 | 사실 외 프레임 미처리 |
| 프리벙킹 | 예방 | 조작기법 사전학습 | 적시성·반복교육 | 과잉경계·효과지속성 |
| 미디어리터러시 | 역량 | 교육·검증습관 | 장기 투자 | 효과의 느린 축적 |

# 부록 B. AI 생성 시각 인포그래픽

아래 이미지는 본 보고서의 주제를 시각적으로 요약하기 위해 생성한 개념적 인포그래픽이다. 이미지 안의 작은 텍스트는 시각적 요소로 취급하며, 사실관계와 정의는 본문 및 참고문헌의 서술을 기준으로 한다.


그림 B-1. 대항선전의 개념·역사·전략·심리·윤리·평가를 한 장에 시각화한 AI 생성 인포그래픽

# 참고문헌 및 주요 자료


[1] Garth S. Jowett & Victoria O'Donnell, Propaganda & Persuasion, 7th ed., SAGE, 2019.

[2] Jacques Ellul, Propaganda: The Formation of Men's Attitudes, Vintage, 1965.

[3] Harold D. Lasswell, Propaganda Technique in the World War, 1927.

[4] Edward Bernays, Propaganda, 1928.

[5] Walter Lippmann, Public Opinion, 1922.

[6] Carl I. Hovland, Irving L. Janis & Harold H. Kelley, Communication and Persuasion, Yale University Press, 1953.

[7] Leon Festinger, A Theory of Cognitive Dissonance, Stanford University Press, 1957.

[8] William J. McGuire, "The Effectiveness of Supportive and Refutational Defenses in Immunizing and Restoring Beliefs Against Persuasion," Sociometry 24(2), 1961.

[9] Daniel Kahneman & Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," Econometrica 47(2), 1979.

[10] Paul M. A. Linebarger, Psychological Warfare, 1948. Project Gutenberg edition. https://www.gutenberg.org/files/48612/48612-h/48612-h.htm

[11] Christopher Paul & Miriam Matthews, The Russian "Firehose of Falsehood" Propaganda Model, RAND, 2016. https://www.rand.org/pubs/perspectives/PE198.html

[12] Todd C. Helmus et al., Understanding Russian Propaganda in Eastern Europe, RAND, 2018. https://www.rand.org/pubs/research_reports/RR2237.html

[13] Elizabeth Bodine-Baron et al., Countering Russian Social Media Influence, RAND, 2018. https://www.rand.org/pubs/research_reports/RR2740.html

[14] Raphael S. Cohen et al., Combating Foreign Disinformation on Social Media, RAND, 2021. https://www.rand.org/pubs/research_reports/RR4373z1.html

[15] Colin P. Clarke/McCulloch et al., Evaluating the Effectiveness of Public Communication Campaigns and Their Implications for Strategic Competition, RAND, 2021. https://www.rand.org/pubs/research_reports/RRA412-2.html

[16] Chenguang Lu et al., "Psychological Inoculation for Credibility Assessment, Sharing Intention, and Discernment of Misinformation," meta-analysis, 2023. https://pmc.ncbi.nlm.nih.gov/articles/PMC10498317/

[17] Briony Swire-Thompson et al., "The backfire effect after correcting misinformation is strongly associated with reliability," Journal of Experimental Psychology: General, 2022. https://pmc.ncbi.nlm.nih.gov/articles/PMC9283209/

[18] Sander van der Linden, "Countering misinformation through psychological inoculation," Advances in Experimental Social Psychology 69, 2024. https://www.sdmlab.psychol.cam.ac.uk/files/media/countering.pdf

[19] David M. J. Lazer et al., "The science of fake news," Science 359(6380), 2018.

[20] UNESCO, Journalism, "Fake News" & Disinformation: Handbook for Journalism Education and Training, 2018. https://www.unesco.org/en/articles/journalism-fake-news-disinformation

[21] OECD, Facts not Fakes: Tackling Disinformation, Strengthening Information Integrity, 2024. https://www.oecd.org/en/publications/facts-not-fakes-tackling-disinformation-strengthening-information-integrity_d909ff7a-en.html

[22] NIST, Reducing Risks Posed by Synthetic Content, NIST AI 100-4, 2024. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-4.pdf

[23] CISA, Risk in Focus: Generative A.I. and the 2024 Election Cycle, 2024. https://www.cisa.gov/resources-tools/resources/risk-focus-generative-ai-and-2024-election-cycle

[24] NATO Allied Command Transformation, Cognitive Warfare. https://www.act.nato.int/activities/cognitive-warfare/

[25] NATO Science & Technology Organization, Chief Scientist Research Report: Cognitive Warfare. https://www.sto.nato.int/wp-content/uploads/chief-scientist-report-cognitive-warfare-final.pdf

[26] Hedayah & ICCT, Developing Effective Counter-Narrative Frameworks for Countering Violent Extremism, 2014. https://icct.nl/sites/default/files/2023-01/Developing%20Effective%20CN%20Frameworks_Hedayah_ICCT_Report_FINAL.pdf

[27] ICCT, "Don't Just Counter-Message; Counter-Engage," 2018. https://icct.nl/publication/dont-just-counter-message-counter-engage

[28] U.S. Department of State, Foreign Affairs Manual: Global Engagement Center - mission and authorities (historical). https://fam.state.gov/FAM/10FAM/10FAM0510.html

[29] U.S. Department of State, Framework to Counter Foreign State Information Manipulation, 2024 (archived). https://2021-2025.state.gov/the-framework-to-counter-foreign-state-information-manipulation/

[30] OHCHR, International Covenant on Civil and Political Rights, Articles 19-20; General Comment No. 34 framework. https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights

[31] European Commission, Digital Services Act - platform systemic risk and electoral integrity framework. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act

[32] Radio Free Europe/Radio Liberty, RFE and the Hungarian Revolution - original broadcasts and retrospective research. https://about.rferl.org/article/rfe-and-the-hungarian-revolution-original-broadcasts-reporters-diary-now-online/

[33] National Security Archive, The 1956 Hungarian Revolution - declassified documents and analysis. https://nsarchive2.gwu.edu/NSAEBB/NSAEBB76/

**[34]** Alberto M. Fernandez, Here to Stay and Growing: Combating ISIS Propaganda Networks, Brookings, 2015. https://www.brookings.edu/wp-content/uploads/2016/07/IS-Propaganda_Web_English_v2-1.pdf

**[35]** RAND, Practical Terrorism Prevention: Reexamining U.S. National Approaches to Addressing the Threat of Ideologically Motivated Violence, 2019. https://www.rand.org/pubs/research_reports/RR2647.html

**[36]** Peter Pomerantsev, How to Win an Information War: The Propagandist Who Outwitted Hitler, 2024.

**[37]** NIST, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST AI 600-1, 2024. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

**[38]** ICCT, Exploitation of Generative AI by Terrorist Groups, 2024. https://icct.nl/publication/exploitation-generative-ai-terrorist-groups

**[39]** RAND, A Compendium of Recommendations for Countering Russian and Other State-Sponsored Propaganda, 2021. https://www.rand.org/pubs/research_reports/RRA894-1.html

**[40]** G7 Rapid Response Mechanism, Joint Statement on a Russian Influence Campaign, 2025. https://2021-2025.state.gov/office-of-the-spokesperson/releases/2025/01/joint-statement-by-the-g7-rapid-response-mechanism-rrm-on-a-russian-influence-campaign

**[41]** U.S. Joint Staff training/doctrine material, “counterpropaganda: Products and actions designed to nullify propaganda or mitigate its effects,” 1999. https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/Joint_Staff/16-F-1727_CJCSM_3500.08_26-May-1999.pdf

## 사실·해석 구분에 관한 주석

본 보고서는 확인 가능한 역사적 사실(기관 존재, 문서, 방송, 법규, 연구결과)과 학계·정책분야의 해석(효과크기, 인과관계, “성공/실패” 평가)을 구분하려 했다. 특히 1956 년 RFE 의 영향, 개별 국가의 정보작전 의도, 심리적 backfire 효과의 보편성, 생성형 AI 가 실제 정치행동에 미치는 순효과는 학계에서 논쟁이 지속되는 영역이다. 따라서 단정적 인과 표현보다 “가능성, 조건, 증거수준”을 표시하는 것이 적절하다.