흑색선전(黑色宣傳, Black Propaganda)

출처·정체성·신뢰를 조작하여 인식과 행동을 변화시키는 정보전 메커니즘

AI 생성 개념 삽화: 출처가 위장된 정보의 상징

작성자: 코리아베스트 https://koreabest.org

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

2026 년 8 월

연구·교육·탐지·방어 목적의 보고서입니다. 실제 기만 캠페인의 설계, 표적화, 계정망 구축, 은폐, 확산 최적화 등 실행 매뉴얼은 제공하지 않습니다.

요약: 흑색선전의 핵심은 "거짓 내용"보다 "거짓 출처"다

흑색선전은 실제 발신자를 숨기는 데서 그치지 않고, 메시지가 적대 세력·제 3 자·독립 언론·자발적 시민·내부고발자 등 다른 주체에게서 나온 것처럼 잘못 귀속되도록 설계되는 선전이다. 따라서 분류의 핵심 축은 content falsity(내용의 거짓)와 source falsification/source deception(출처의 거짓)을 분리하는 데 있다. 고전적 백색·회색·흑색 분류는 무엇보다 발신자와 출처의 식별·귀속 상태를 가리키며, 사실성은 별도의 축이다. 흑색선전은 진실한 사실을 사용하면서도 출처를 위장할 수 있고, 반대로 공개된 정부가 거짓말을 해도 그것만으로 자동으로 흑색선전이 되는 것은 아니다. [R1][R3][R7]

효과의 원천은 "내가 이 메시지를 누구의 말이라고 생각하는가"에 있다. 수용자는 내용만 평가하지 않는다. 집단 정체성, 전문성, 내부자성, 독립성, 사회적 증거, 친숙성, 반복 노출, 감정적 각성 등을 통해 출처의 신뢰도를 추정한다. 흑색선전은 이 신뢰 추정 장치를 공격한다. 그래서 가장 강력한 형태는 거짓말만 반복하는 것이 아니라, 정확한 정보와 내부자적 세부사항을 섞어 신뢰를 만든 다음, 결정적 순간에 왜곡된 주장·루머·행동 유도 신호를 삽입하는 방식이다. 다만 실제 효과는 캠페인마다 크게 다르고, 온라인 영향작전이 거대해 보여도 진짜 이용자에게 거의 도달하지 못한 사례가 적지 않다. [R13][R17][R18]

디지털 시대에는 가짜 페르소나, 위장 뉴스사이트, 봇·시빌 계정, AI 생성 이미지·음성·영상, 인플루언서, 온라인 커뮤니티, 번역·재전파, 전통 언론의 재인용이 연결되면서 출처 세탁(source laundering)이 빨라진다. 이때 "어디서 봤는지"가 기억에서 사라지고 "여러 곳에서 본 이야기"라는 친숙성만 남으면, 반복 노출 자체가 신빙성을 높일 수 있다. 반면 허위정보가 언제나 봇 때문에 퍼지는 것은 아니다. 대규모 트위터 연구는 거짓 뉴스가 더 빠르고 넓게 확산됐지만, 봇은 참·거짓을 모두 비슷하게 가속했고 인간의 공유 행동이 차이를 설명하는 핵심이었다고 보고했다. [R23][R24]

방어의 중심은 단순 팩트체크보다 "출처 포렌식"이다. 최초 게시물, 발신자 실체, 계정 생성·행동 패턴, 원본 파일, 문서·이미지 메타데이터, 도메인·인프라 관계, 동일 메시지의 동시 반복, 교차 플랫폼 확산 순서, 독립적 확인원, 전략적 수혜자를 함께 본다. 또한 "내가 싫어하는 주장 = 선전", "익명 = 정보기관", "거짓 내용 = 흑색선전"이라는 세 가지 오류를 피해야 한다. attribution 은 확률적 판단이며, 기술적 유사성·정치적 이해관계만으로 국가 행위자를 단정해서는 안 된다.

목차

1. 정의와 어원: 선전의 "색"은 도덕성이 아니라 출처 귀속을 뜻한다

1.1 propaganda 와 black propaganda

propaganda 의 어원은 라틴어 propagare(퍼뜨리다, 번식시키다)와 1622 년 교황청의 Congregatio de Propaganda Fide 에서 자주 설명된다. 현대적 의미의 선전은 특정 정치·사회·군사적 목적을 위해 상징, 사실, 해석, 감정, 정체성, 매체를 선택적으로 조직하여 인식·태도·행동에 영향을 미치려는 체계적 커뮤니케이션을 말한다. 선전은 반드시 거짓일 필요가 없으며, 사실의 선택·강조·프레이밍만으로도 선전이 될 수 있다. [R1][R2][R4]

black propaganda 는 이 가운데 실제 발신자나 후원자를 허위로 귀속시키는 유형이다. 예컨대 A 국 정보기관이 만든 메시지를 B 국 군인 내부의 반체제 조직이 만든 것처럼 꾸미거나, 외국 정부가 만든 웹사이트를 현지 독립 언론처럼 보이게 하는 경우가 전형적이다. CIA 의 공개 설명도 흑색선전을 "한쪽이 만든 것처럼 보이지만 실제로는 반대편이 만든 정보"로 요약한다. [R7]

1.2 백색·회색·흑색선전 비교

유형	출처의 상태	정보의 진위	기만 수준	주요 목적
일반 선전	실제·불명·위장 모두 가능	참·거짓·혼합 모두 가능	필수 아님	설득·동원·정당화·억제
백색선전	출처가 실제 발신자와 일치	대체로 사실·선택적 사실·해석	낮음	공개적 설득, 신뢰 구축
회색선전	출처가 불명확·비공개·모호	참·거짓·혼합 가능	중간	책임 회피, 모호성 활용
흑색선전	출처가 의도적으로 잘못 귀속됨	참·거짓·혼합 모두 가능	높음	신뢰 탈취, 분열, 사기 저하, 행동 유도

중요: "흑색 = 거짓 내용"이라고만 정의하면 핵심을 놓친다. 색상 분류의 우선 기준은 source attribution 이다. 흑색선전은 사실을 말하면서도 출처를 위장할 수 있다.

1.3 content falsity 와 source falsification

내용의 거짓(content falsity)은 주장 자체가 사실과 다른지를 묻는다. 출처의 거짓(source falsification)은 "누가 이 메시지를 만들었는가·후원했는가·전달했는가"에 관한 표시가 실제와 다른지를 묻는다. 두 축은 독립적이다. 사실인 문서를 위조된 내부고발자 계정이 유출하면 내용은 참일 수 있지만 출처는 거짓일 수 있다. 반대로 정부 대변인이 공개 브리핑에서 허위 주장을 하면 내용은 거짓이지만 출처는 공개되어 있으므로 전형적 흑색선전은 아니다.

흑색선전에서 source falsification 이 중요한 이유는 수용자가 메시지의 의미와 신뢰를 출처에 의존해 해석하기 때문이다. "적 내부의 불만", "독립 언론의 폭로", "평범한 현지 시민의 자발적 의견"처럼 보이게 되면 같은 문장도 전혀 다른 설득력을 가질 수 있다. 출처 위장은 메시지의 사회적 맥락 자체를 위조하는 행위다.

AI 생성 개념 삽화: 숨겨진 발신자와 위장된 행위자

2. 관련 개념 지도: 서로 겹치지만 같은 개념은 아니다

개념	구분 핵심
Propaganda	목표 집단의 태도·행동에 영향을 주기 위해 메시지와 상징을 조직하는 넓은 범주. 진실 여부와 출처 공개 여부는 다양하다.
Disinformation	기만 의도를 가지고 의도적으로 허위·오도 정보를 생산·유통하는 것. 중심 축은 내용과 의도다. [R12]

개념	구분 핵심
Misinformation	거짓·부정확 정보가 기만·해악 의도 없이 공유되는 것. [R12]
Malinformation	진짜 정보가 맥락 왜곡, 선택적 공개, 사생활 침해 등 해로운 방식으로 이용되는 것. [R12]
Psychological warfare	전쟁·경쟁 상황에서 적·중립·우호 집단의 심리와 행동을 겨냥하는 전략적 활동의 역사적 상위 개념.
PSYOP/MISO	선택된 정보를 외국 목표집단에 전달해 감정·동기·추론·행동에 영향을 미치는 계획된 군사 활동. 용어·교리는 시대별로 변한다.
Information operations	정보환경에서 정보 관련 능력을 통합해 의사결정·행동·시스템에 영향을 주거나 보호하는 더 넓은 작전 범주.
Influence operations	국가·비국가 행위자가 타인의 판단·행동을 바꾸려는 다양한 수단을 묶는 실무적 포괄어.
Perception management	상대가 현실을 해석하는 방식을 바꾸려는 광범위한 개념. 냉전기 미국 문헌에서 자주 사용.
Deception	상대의 판단을 오도하기 위해 의도적으로 잘못된 단서·표상·신호를 제공하는 행위.
Strategic communication	정책·행동·메시지를 조정해 전략적 목표를 지원하는 커뮤니케이션. 이상적으로는 공개성과 신뢰성을 중시.
Fake news	학술적으로 지나치게 모호하고 정치적 낙인으로도 사용되어 분석 용어로는 주의가 필요. [R12]
Rumor	공식적으로 확인되지 않은 채 사회적으로 순환하는 주장. 거짓일 수도 참일 수도 있다.
Conspiracy theory	중요 사건을 비밀 공모로 설명하는 서사. 존재 자체가 곧 거짓은 아니지만 반증 불가능 구조를 취할 수 있다.
Astroturfing	조직된 후원·동원을 숨기고 자발적 풀뿌리 지지나 반대처럼 보이게 하는 행위.
False flag communication	메시지의 발신자·정체성을 상대편 또는 제 3 자로 거짓 표시하는 커뮤니케이션상의 false attribution.
Front organization	실제 후원자·통제자를 숨기고 독립 조직처럼 활동하는 명목상 조직.
Covert influence operation	후원·지휘·목표를 은폐한 영향력 작전. 반드시 타인의 신원을 사칭하지는 않으므로 흑색선전보다 넓다.

2.1 disinformation 과 black propaganda 의 관계

둘은 크게 겹치지만 동일하지 않다. disinformation 은 의도적 허위·오도라는 내용·의도 축에 초점을 맞춘다. black propaganda 는 발신자·출처의 허위 귀속이라는 attribution 축에 초점을 맞춘다. 따라서 "거짓 내용 + 진짜 출처"는 disinformation 이지만 black propaganda 가 아닐 수 있고, "진짜 내용 + 가짜 출처"는 black propaganda 이지만 좁은 의미의 disinformation 정의에는 완전히 들어맞지 않을 수 있다.

상황	Disinformation	Black propaganda
정부가 공개 계정으로 허위 통계를 발표	가능	보통 아님
외국 정보기관이 현지 시민 계정으로 거짓 루머 유포	예	예
실제 문서를 가짜 내부고발자 계정이 공개	정의에 따라 일부/아님	예
출처 불명 사이트가 혼합 정보를 유통	가능	회색선전에 더 가까울 수 있음

3. 흑색선전의 구조: 신뢰를 탈취하는 인과 사슬

분석 단위는 개별 게시물이 아니라 작전 체계다. 다음의 연쇄를 따라가면 "무엇이 거짓인가"보다 "어디에서 신뢰가 조작되는가"를 볼 수 있다.

단계	분석 질문
1 실제 발신자	정부·정보기관·군·정당·기업·이익집단·비국가 조직 등 실제 기획·후원 주체.
2 위장된 발신자	적대 세력 내부자, 독립 언론, NGO, 시민운동가, 전문가, 지역 주민, 내부고발자, 제 3 국 조직 등으로 가장.
3 메시지	사실, 반쪽 진실, 조작 문서, 루머, 이미지, 밈, 음성, 영상, 선택적 유출 등이 혼합될 수 있음.
4 전달매체	비밀 라디오, 전단, 우편, 위장 신문, 웹사이트, SNS, 메신저, 팟캐스트, 영상 플랫폼.
5 목표집단	적군, 점령지 주민, 유권자, 특정 인종·종교·세대·지역, 투자자, 언론인, 정책결정자.
6 심리적 취약점	불안, 분노, 수치, 정체성 갈등, 불신, 정보 부족, 권위 의존, 사회적 고립.
7 확산 네트워크	봇·가짜 계정, 진짜 지지자, 인플루언서, 커뮤니티, 제 3 자 언론, 전통 매체 재인용.
8 인식·행동 변화	신뢰 약화, 후보 지지 변화, 투표 포기, 사기 저하, 탈영, 시위, 매도·매수, 정책 압박 등.
9 정치·군사·사회적 효과	분열, 의사결정 지연, 자원 소모, 동맹 불신, 선거 정당성 훼손, 시장 변동, 장기적 냉소주의.

효과 측정에서는 노출(exposure), 참여(engagement), 믿음(belief), 행동(behavior), 최종 결과(outcome)를 분리해야 한다. 조회수나 계정 수만으로 선거 결과·전쟁 결과를 바꿨다고 결론 내리면 안 된다.

3.1 신뢰를 만드는 장치

- 정확한 세부정보와 실제 뉴스 일부를 먼저 제공해 "내부자성"을 연출
- 목표집단의 방언·문화·직업 언어를 사용해 동일집단 신호를 제공
- 권위 있는 로고·서식·도메인·문서 형식을 모방
- 여러 독립 계정에서 같은 메시지가 나온 것처럼 보여 사회적 증거를 조작
- 진짜 언론이 재인용하게 만들어 원래 출처를 잊게 함
- 사실 확인이 어려운 위기·전쟁·속보 상황을 이용

4. 역사적 발전: 라디오의 위장 발신자에서 AI 페르소나까지

4.1 제 1 차 세계대전: 대중선전의 산업화

제 1 차 세계대전은 국가가 신문, 포스터, 전단, 전신, 영화, 검열을 통합해 대중 심리를 조직적으로 관리한 전환점이었다. 전쟁 잔혹행위 보도와 적 이미지의 과장, 중립국 여론전, 전선의 전단전이 확대되면서 "사실을 얼마나 왜곡했는가"뿐 아니라 "누가 만든 메시지인가"를 숨기는 기술이 발전했다. 다만 오늘날의 백색·회색·흑색 구분을 당시 모든 사례에 소급 적용할 때는 주의가 필요하다. [R4]

4.2 제 2 차 세계대전: 흑색 라디오의 전성기

영국 Political Warfare Executive(PWE)와 미국 OSS Morale Operations 는 적국 내부에서 나온 것처럼 보이는 라디오, 문서, 루머, 전단을 제작했다. Sefton Delmer 가 관여한 Gustav Siegfried Eins 와 Soldatensender Calais 는 독일 내부의 불만 세력 또는 군 방송처럼 들리게 설계된 대표 사례다. 핵심은 모든 내용을 거짓으로 만드는 것이 아니라, 실제 전황·지역 정보·군인들의 관심사를 섞어 신뢰를 확보한 뒤 나치 지도부의 부패·무능·불공정을 부각하는 것이었다. OSS 도 독일군 내부에서 나온 것처럼 보이는 메시지와 루머를 활용했다. CIA 의 역사 자료는 OSS Morale Operations 가 "black or subversive propaganda"를 사용했다고 기록한다. [R5][R6][R8][R9]

4.3 냉전: covert sponsorship 과 black attribution 의 분화

냉전기의 비밀 영향공작은 흑색선전과 회색선전, 프런트 조직, 언론 보조금, 문화조직, 위조문서, 루머를 혼합했다. 중요한 구분은 "후원자가 비공개"인 것과 "다른 주체가 만든 것처럼 사칭"하는 것이 다르다는 점이다. 예를 들어 초기 Radio Free Europe/Radio Liberty 의 CIA 후원 은폐는 covert sponsorship/gray 영역으로 분류하는 편이 더 정확하며, 방송국 자체가

소련 정부 방송을 사칭한 것은 아니었다. 반면 KGB 의 위조문서, 제 3 국 언론을 통한 허위 이야기 식재, 가짜 미국 기관 문서는 흑색선전의 전형적 성격을 가진다. [R37]

4.4 한국전쟁: 생물전 주장과 "논쟁적 귀속"의 교훈

1951-1952 년 북한·중국·소련은 미국의 생물무기 사용을 주장하며 국제적 선전전을 전개했다. Milton Leitenberg 와 Kathryn Weathersby 등은 1990 년대 공개된 소련권 자료를 근거로 이 주장이 조작된 캠페인이었다고 평가한다. 반면 Stephen Endicott 와 Edward Hagerman 은 해당 소련 문서의 출처·검증 가능성과 해석에 이의를 제기하며 미국의 생물전 가능성을 완전히 배제할 수 없다고 주장했다. 오늘날 다수 서구 연구는 조작설에 무게를 두지만, 이 사안은 "확인된 사실"과 "학술적 해석"을 분리해 표기해야 하는 대표 사례다. [R31][R32][R33]

4.5 베트남전: 공식문서로 확인되는 black radio 와 black letters

미 국무부 Foreign Relations of the United States(FRUS)의 1970 년·1972 년 문서는 CIA 가 북베트남과 남베트남 민족해방전선을 대상으로 black radio, black letter mailing, black leaflets, fabricated enemy documents 등을 운용했다고 명시한다. 이는 "흑색선전"을 추상 개념이 아니라 당시 정책 문서의 실제 분류로 확인할 수 있는 자료다. 해당 보고서는 청취와 적측의 경계 반응을 성과 지표로 제시했지만, 개별 행동 변화가 방송 때문인지 다른 전황 요인 때문인지는 분리하기 어렵다. [R10][R11]

4.6 인터넷·SNS 시대: 가짜 페르소나와 위장 매체

2010 년대 이후 흑색선전의 주요 변화는 발신자 위장이 대규모 계정·웹사이트 네트워크로 자동화·분산되었다는 점이다. 러시아 Internet Research Agency(IRA)는 미국 시민처럼 보이는 소셜미디어 계정을 만들고 정치적 집회와 논쟁을 촉진한 혐의로 미국 법무부에 기소됐다. Secondary Infektion 은 가짜 계정과 위조문서를 300 개가 넘는 플랫폼에 흩뿌렸지만 실제 영향력은 대체로 낮았다. 2024 년 DOJ 가 공개한 Doppelganger 사건은 러시아 정부 지시를 받은 것으로 주장되는 네트워크가 미국·유럽의 실제 언론 도메인을 흉내 낸 사이트, 가짜 시민 계정, 인플루언서, 일부 AI 생성 광고를 결합했다고 설명한다. [R14][R17][R15]

중국 연계 Spamouflage 도 수천 개 가짜 계정과 다수 플랫폼을 활용해 친중·반서방 메시지를 확산했으며 Meta 는 2023 년 이를 중국 법집행기관과 연계된 인물들에게 연결했다. 그러나 Meta 는 이 네트워크가 대규모임에도 진짜 이용자에게 도달하는 데 지속적으로 어려움을 겪었다고 평가했다. 규모와 효과는 동일하지 않다. [R18][R19]

AI 생성 개념 삽화: 국경을 넘는 정보 네트워크

5. 대표적 역사·현대 사례 10 개

아래 표는 "실제 발신자 - 가장된 발신자 - 목표집단 - 메시지 - 전달수단 - 신뢰 장치 - 효과 - 발각 과정 - 성공·실패 원인"의 동일 기준으로 비교한다. 효과 평가는 직접 인과가 확인된 경우와 정황적·논쟁적 평가를 구분했다.

사례	실제 발신자	가장된 발신자	목표집단	메시지	수단	신뢰 장치	실제 효과	발각	평가
1. 1782 Benjamin Franklin 의 가짜 신문 보충판	Franklin/미 독립 측	보스턴 신문·영국 측 자료처럼 보이는 형식	영국·유럽 여론	원주민 두피·영국 잔혹행위 관련 조작 서사	인쇄물	신문 형식·문서 외관	정량효과 불명	후대 문서·연구	초기 black propaganda 의 고전적 선례로 자주 인용
2. Gustav Siegfried Eins	영국 PWE	독일군 내부의 강경 보수 반체제 방송	독일 군·민간	나치 지도부의 부패·무능 비판	비밀 라디오	독일어·군대 문화·내부자 정보	독일 당국의 우려 기록, 행동효과는 불확실	전후 PWE 기록	문화적 정교함과 사실 혼합
3. Soldatensender Calais	영국 PWE	독일군 방송처럼 위장	독일군	실제 뉴스와 사기 저하 정보 혼합	고출력 라디오	빠른 뉴스·지역 세부정보·음악	청취·나치 우려 자료 존재, 전략 효과 분리 어려움	전후 문서·회고록	신뢰성 있는 진실을 "커버"로 사용
4. OSS Operation Sauerkraut	미 OSS Morale Operations	독일군·독일인 내부 메시지	독일군	지도부 불화·패배 전망·귀환 유도	전단·루머·가짜 문서	독일 군인·포로 활용, 현지 문체	전술적 효과 일부 보고, 전체 영향 불명	OSS/CIA 역사 공개	현지성·내부자성
5. 1954 Guatemala Voice of Liberation	CIA 지원 네트워크	과테말라 국내 반정부 비밀방송	과테말라 군·시민	반정부 세력 규모 과장, 정부 약화 서사	라디오	현지 방송처럼 보이는 위치·언어·속보	군·사회 불안에 기여했다는 역사 평가, 단독 인과는 불가	CIA 문서·후대 연구	정부 방송 공백과 정치·군사 압박이 결합
6. 한국전쟁 생물전 주장	북한·중국·소련 측으로 귀속된 국제 캠페인	피해 증언·과학 조사·공식 주장	한국·중국·국제여론	미국의 생물무기 사용 주장	언론·국제기구·자료	전쟁 공포·과학 권위	국제적 공명 큼. 사실성·조작 여부는 역사 논쟁	소련권 문서·반론 연구	위기·검증 제한; 자료 출처 논쟁 때문에 신중한 평가 필요
7. 베트남전 CIA black radio/letters	CIA/ARVN 협력	북베트남·NLF 내부 소식·반대파처럼 위장	북베트남·남베트남 대상	체제 불신·사기 저하·전향 유도	라디오·우편·전단·가짜 적 문서	내부자 형식·반복 방송	FRUS 는 청취·적 반응을 성과로 제시; 인과 분리 어려움	기밀해제 FRUS	지속성은 높았으나 전쟁 결과와 직접 연결 불가
8. Soviet AIDS/Operation Denver-INFEKTION	소련 KGB active measures 로 널리 평가	제 3 국 언론·과학자·문서	개도국·서방 여론	AIDS 가 미군 생물무기 연구에서 발생했다는 주장	언론·위조자료·재전파	제 3 국 출처와 과학 언어	장기 생존한 음모서사; 정확한 행동효과는 불명	동구권 문서·미 정부 조사·연구	source laundering 과 기존 반미 불신 활용
9. Russia IRA 2014-2018	러시아 IRA 로 DOJ 기소	미국인·미국 정치단체로 위장한 계정	미국 유권자·활동가	갈등 이슈·후보 관련 메시지	Facebook/ Instagram/ Twitter 등	미국식 페르소나·현지 서버·신분자료	상당한 노출·참여 확인; 2016 결과 변경은 DOJ 가 주장하지 않음	플랫폼·수사·기소	진짜 사회갈등에 편승, 효과 크기는 논쟁적
10. Secondary Infektion / Doppelganger / Spamouflage	러시아 기반·정부 연계 주장, 중국 연계 네트워크	가짜 시민·위장 언론·조작 문서	미국·유럽·아시아 등	분열·반서방·친러/친중 서사	다중 플랫폼·위장 도메인·AI 일부	언론 외관·다중 출처·가짜 사회적 증거	많은 작전은 진짜 참여가 낮았음; Doppelganger 는 2024 DOJ 조치	플랫폼·OSINT·법집행	규모는 크지만 신뢰·현지화 부족 시 실패

사례 6 은 의도적으로 "논쟁적"으로 표시했다. 역사적 정보전 연구에서 가장 위험한 오류는 현재의 정치적 입장에 맞춰 귀속을 확정하는 것이다.

5.1 사례별 상세 분석

사례 1. Benjamin Franklin 의 1782 년 가짜 신문 보충판 - 초기 선례

분석 항목	내용
실제 발신자	Benjamin Franklin 과 미국 독립 측의 정보·외교전
가장된 발신자	실제 북미 신문 또는 영국 측에서 흘러나온 자료처럼 보이는 인쇄물
목표집단	영국과 유럽의 여론 및 정책 엘리트
메시지	영국이 동맹 원주민의 잔혹행위를 조장한다는 충격적 서사. 문서의 사실성보다 신문 형식을 이용한 허위 귀속이 핵심이었다.
전달수단	인쇄물·재인용
신뢰 장치	당시 신문 문체와 문서 형식, 구체적 숫자·증언을 흉내 낸 세부정보
실제 효과	정량적으로 입증하기 어렵다. 후대 연구에서 초기 black propaganda 의 상징적 사례로 널리 인용된다.
발각 과정	Franklin 문서와 후대 역사연구를 통해 제작 경위가 알려짐
성공·실패 원인	인쇄 언론에 대한 신뢰와 잔혹행위 서사의 감정성이 강점이었지만 실제 정책 효과는 분리하기 어렵다.

사례 2. Gustav Siegfried Eins - "적 내부의 목소리"를 가장한 PWE 방송

분석 항목	내용
실제 발신자	영국 Political Warfare Executive 계열, Sefton Delmer 팀
가장된 발신자	나치에 충성하지만 당 지도부의 부패를 혐오하는 독일 내부 강경파·군인 집단
목표집단	독일군과 독일 민간 청취자
메시지	지도부의 부패·성적 스캔들·특권·무능을 공격하며 "진짜 독일"과 당 엘리트를 분리
전달수단	비밀 단파 라디오
신뢰 장치	거친 내부자 어조, 군대·지역 세부정보, 실제 뉴스와 소문 혼합
실제 효과	독일 당국의 관심과 대응은 기록되지만 사기 저하·탈영을 방송 단독 효과로 계량하기는 어렵다.
발각 과정	전후 PWE 기록과 Delmer 회고록, 후대 연구
성공·실패 원인	정교한 문화적 위장과 실제 정보 혼합이 강점. 과도한 자극은 진위 의심을 키울 위험이 있었다. [R5][R6]

사례 3. Soldatensender Calais - 진실을 많이 말한 흑색 라디오

분석 항목	내용
실제 발신자	영국 PWE
가장된 발신자	독일군을 위한 국내·전선 방송
목표집단	독일군 병사와 가족
메시지	빠른 전황·음악·군대 소식에 지도부 비판과 사기 저하 소재를 삽입
전달수단	고출력 라디오
신뢰 장치	공식 독일 방송보다 빠른 실제 뉴스, 구체적 부대·지역 정보, 익숙한 음악
실제 효과	청취와 독일 측 우려의 정황은 있으나 전쟁 결과에 미친 독립 효과는 확인 곤란
발각 과정	전후 영국 문서·회고록
성공·실패 원인	"거짓을 믿게 하려면 먼저 믿을 만한 진실을 제공한다"는 흑색선전의 전형을 보여준다. [R5][R6]

사례 4. OSS Operation Sauerkraut - 적군 내부 메시지의 위장

분석 항목	내용
실제 발신자	미국 OSS Morale Operations
가장된 발신자	독일군 내부자·독일 시민·현지 루머
목표집단	독일군 병사
메시지	패전 전망, 나치 지도부에 대한 불신, 안전한 귀환·항복 유도 등
전달수단	전단·루머·가짜 문서·현장 유포
신뢰 장치	독일 포로·탈영병의 언어·군대 경험을 활용한 현지성
실제 효과	일부 전술적 성과 보고가 있으나 전체 사기 붕괴에 대한 기여도는 불분명
발각 과정	OSS 기록과 CIA 역사 공개
성공·실패 원인	목표집단 문화에 맞춘 언어가 강점. 전선 상황 자체가 설득력을 좌우했다. [R8]

사례 5. 1954 Guatemala "Voice of Liberation" - 반군 규모를 크게 보이게 한 비밀방송

분석 항목	내용
실제 발신자	CIA 가 지원한 PBSUCCESS 심리전 네트워크
가장된 발신자	과테말라 국내 반정부·해방 세력의 방송
목표집단	과테말라 군·정부 관계자·시민
메시지	반군의 존재와 군사적 진전을 실제보다 크게 보이게 하고 정부의 통제력에 의문을 제기
전달수단	라디오와 기타 심리전 수단

분석 항목	내용
신뢰 장치	현지 언어·속보 형식, 실제 군사 압박과 결합된 정보 공백
실제 효과	쿠데타 환경의 불안·사기 저하에 기여했다는 평가가 많지만, 외교·군사·경제 압박과 분리해 방송의 독립효과를 산정하기 어렵다.
발각 과정	CIA 기밀해제 문서와 Nick Cullather 등 후대 역사연구
성공·실패 원인	정보전이 실제 군사·정치 압박과 동시에 작동할 때 신빙성이 강화된 사례.

사례 6. 한국전쟁 생물전 주장 - 흑색선전의 "경계 사례"

분석 항목	내용
실제 발신자	북한·중국·소련의 공식·비공식 선전 체계로 보는 연구가 다수
가장된 발신자	공식 주장 자체는 출처가 공개되어 있어 전형적 black propaganda 는 아니다. 다만 조작된 증거·증언이 있었다면 source deception 요소가 생긴다.
목표집단	한국·중국 주민, 공산권·비동맹권·서방 국제여론
메시지	미국이 세균·생물무기를 사용했다는 주장
전달수단	언론, 공식 성명, 국제 조사, 사진·증언·문서
신뢰 장치	과학자·조사단의 권위, 전쟁 질병·공포, 미국의 실제 생물무기 연구 역사와 결합
실제 효과	당시 국제적 공명이 컸다. 그러나 사실성은 오랫동안 논쟁됐고, 현재도 일부 연구자는 반대 해석을 제시한다.
발각 과정	1990 년대 소련권 자료 공개와 Leitenberg/Weathersby 분석, Endicott/Hagerman 의 반론
성공·실패 원인	검증이 어려운 전쟁 환경이 확산에 유리했다. 동시에 사료의 provenance 자체가 논쟁 대상이 되어 "귀속 포렌식"의 중요성을 보여준다. [R31][R32][R33]

사례 7. 베트남전 CIA black radio·black letters - 정책 문서로 확인된 작전

분석 항목	내용
실제 발신자	CIA 와 남베트남 파트너
가장된 발신자	북베트남·NLF 내부 반대파 또는 내부 소식통처럼 보이는 발신자
목표집단	북베트남 주민·군, NLF 관련 집단
메시지	체제 내부 불신, 사기 저하, 지도부와 개인의 이해 충돌, 전향 유도
전달수단	black radio, black letters, black leaflets, fabricated enemy documents
신뢰 장치	적 내부 문서·방송의 형식과 언어를 재현
실제 효과	FRUS 는 청취 증가, 귀순자 청취, 적측 경고 등을 성과 신호로 제시한다. 그러나 이는 작전 자체의 보고서이므로 독립적 효과 검증과 구분해야 한다.
발각 과정	기밀해제된 FRUS 40 Committee 문서
성공·실패 원인	장기간 지속됐고 적의 관심을 끈 것은 확인되지만 베트남전의 전략 결과를 설명할 정도의 효과는 입증되지 않았다. [R10][R11]

사례 8. Soviet AIDS disinformation / Operation DENVER-INFEKTION

분석 항목	내용
실제 발신자	소련 KGB active measures 체계로 널리 평가
가장된 발신자	제 3 국 언론·익명 독자·과학적 출처·독립 연구자
목표집단	개도국과 서방의 반미·반군사기지 정서가 강한 청중
메시지	AIDS 가 미국 군사 생물무기 연구에서 만들어졌다는 주장
전달수단	제 3 국 신문 식재, 번역·재인용, 위조·과학적 외관의 자료
신뢰 장치	미국이 아닌 제 3 국에서 시작된 것처럼 보이는 source laundering 과 과학 전문어
실제 효과	주장의 변형이 수십 년간 음모론 생태계에 남았다. 특정 정책·행동 변화의 크기는 측정하기 어렵다.
발각 과정	미국 정부의 active measures 조사, 동구권·소련 자료, 후대 연구
성공·실패 원인	실제 냉전 불신과 식민주의·생물전의 역사적 기억에 접속해 장기 생존했다.

분석 항목	내용
	[R37]

사례 9. Russia Internet Research Agency - 미국 시민을 가장한 정치 참여

분석 항목	내용
실제 발신자	미국 DOJ 가 러시아 Internet Research Agency 및 관련자를 기소
가장된 발신자	미국 시민, 지역 활동가, 정치·사회 운동 페이지
목표집단	미국 유권자·정치활동가·온라인 커뮤니티
메시지	후보 지지·반대뿐 아니라 인종·이민·종교 등 갈등 이슈 전반
전달수단	주요 소셜미디어와 온라인 광고·이벤트 조직
신뢰 장치	미국식 이름·현지 언어·지역 이슈·가짜 조직 정체성
실제 효과	상당한 노출과 실제 이용자의 상호작용은 확인됐다. 그러나 DOJ 기소는 2016 년 선거 결과가 바뀌었다고 결론 내리지 않았다.
발각 과정	플랫폼 조사, 미국 정보·법집행 수사, 2018 년 기소
성공·실패 원인	외부에서 새로운 갈등을 만든 것보다 이미 존재한 미국 사회갈등에 편승한 점이 핵심. [R14]

사례 10. Secondary Infektion·Doppelganger·Spamouflage - 규모와 효과의 분리

분석 항목	내용
실제 발신자	Secondary Infektion 은 러시아 기반의 지속적 행위자로 연구됨; Doppelganger 는 DOJ 가 러시아 정부 지시 네트워크로 주장; Spamouflage 는 Meta 가 중국 법집행기관 연계 인물과 연결
가장된 발신자	가짜 시민, 위장 언론, 가짜 전문가, 위조 문서의 작성자, 독립 온라인 매체
목표집단	미국·유럽·우크라이나·아시아 등 다양한 정치·외교 청중
메시지	동맹 분열, 우크라이나 비방, 서방 기관 불신, 친러·친중 서사 등
전달수단	수백 플랫폼, 위장 도메인, 가짜 계정, 광고, 일부 AI 생성 콘텐츠
신뢰 장치	실제 언론 외관, 다중 언어, 여러 "독립" 출처, 도난·위조 문서의 외관
실제 효과	Secondary Infektion 과 Spamouflage 는 규모에 비해 유기적 참여가 낮았다는 연구·플랫폼 평가가 중요하다.
발각 과정	OSINT 연구, 플랫폼 CIB 조사, 법집행 도메인 압수·기소
성공·실패 원인	대량 생산은 가능했지만 지역 문화의 부자연스러움과 팔로워·신뢰 부족이 확산을 제한했다. [R17][R18][R15]

6. 흑색선전의 심리적 메커니즘

메커니즘	흑색선전에서의 작동 방식
확증편향	기존 신념에 맞는 정보는 더 쉽게 받아들이고 반대 증거는 엄격하게 평가한다. 흑색선전은 목표집단이 이미 믿고 싶은 출처·정체성을 가장한다.
권위편향	전문가·기관·내부자·언론의 권위를 사칭하면 같은 주장도 더 신뢰받을 수 있다.
반복효과·진실착각효과	반복 노출은 친숙성을 높이고 친숙성은 진실성 판단의 휴리스틱이 될 수 있다. 실제 가짜뉴스 헤드라인도 단 한 번의 사전 노출로 정확성 평가가 상승했다. [R24]
집단동조·사회적 증거	여러 계정이 같은 주장을 지지하면 독립적 다수 의견처럼 보인다. 가짜 계정·봇·댓글망은 이 사회적 증거를 인위적으로 만들 수 있다.
내집단·외집단 편향	"우리 집단 사람"처럼 보이는 발신자는 더 신뢰받고, 외집단에 대한 위협 서사는 더 빠르게 결집을 일으킬 수 있다.
공포·분노·혐오	고각성 감정은 공유와 행동을 촉진하고 세밀한 검토를 줄일 수 있다. Vosoughi 등은 거짓 뉴스가 공포·혐오·놀람 반응과 연관됐다고 보고했다. [R23]
도덕적 충격	아동 피해, 배신, 성적 범죄, 국가 모욕 등 도덕적으로 충격적인 소재는 사실 확인 전에 행동을 유도할 수 있다.
정체성 보호	사실 판단이 정체성·소속감과 결합하면 반대 증거가 개인의 사회적 지위를 위

메커니즘	흑색선전에서의 작동 방식
	협하는 것으로 느껴질 수 있다.
음모론적 사고	불확실성과 통제 상실 상황에서 복잡한 사건을 의도적 비밀 공모로 설명하는 서사가 매력적일 수 있다.
정보 과부하	대량 메시지는 상대의 사실확인 능력과 주의력을 소진시키고 "어차피 아무도 모른다"는 냉소주의를 만들 수 있다. RAND 는 러시아식 모델의 고용량·다채널·반복성을 지적한다. [R13]
감정적 전염	온라인 네트워크에서 감정적 언어와 반응이 주변 사용자에게 전파될 수 있다.
루머 확산	공식 정보가 부족하고 사건의 중요성이 높을수록 미확인 정보가 빈 공간을 채우기 쉽다.
적대적 미디어 효과	같은 중립적 보도도 강한 당파성을 가진 집단은 상대편에 편향됐다고 느낄 수 있어, "언론은 모두 거짓"이라는 서사가 강화된다.
출처 기억 소실	시간이 지나면 메시지 내용은 기억하지만 어디서 봤는지는 잊을 수 있다. 이때 원래 낮은 신뢰도의 출처라는 경고가 약화될 수 있다.

7. 흑색선전이 성공하기 쉬운 사회적 조건

조건	왜 취약해지는가
정치적 양극화	상대 진영에 대한 최악의 설명이 이미 준비되어 있어 위장 메시지가 기존 적대감에 접속하기 쉽다.
전쟁·테러·경제위기	시간 압박과 불확실성이 커져 검증보다 속보·감정·소문에 의존한다.
정부·언론에 대한 낮은 신뢰	공식 부인을 오히려 은폐의 증거로 해석할 수 있어 반박이 힘들어진다.
사회갈등·차별 경험	실제 불만과 상처가 존재할수록 외부 행위자가 만든 메시지도 "우리 현실을 아는 내부자의 말"처럼 들릴 수 있다.
검열·정보 부족	독립적 확인원이 적으면 루머·비밀방송·유출문서의 희소가치가 커진다.
언론 생태계 붕괴	지역언론·전문기자의 감소는 출처 검증과 맥락 설명의 중간층을 약화한다.
추천 알고리즘	참·거짓 자체보다 참여 가능성을 최적화하는 시스템은 감정적·새로운 콘텐츠를 확대할 수 있다. 다만 알고리즘이 모든 확산을 설명하는 것은 아니다.
에코체임버·동질 네트워크	서로 비슷한 이용자 사이에서 반복 노출이 독립적 확인처럼 느껴질 수 있다.
인플루언서 네트워크	신뢰가 기관보다 개인에게 이동하면 숨겨진 후원·협찬·외국 연결을 파악하기 어려워질 수 있다.
다언어·번역 환경	원문과 번역문 사이에서 맥락·출처 표기가 손실되며 재전파 과정에서 최초 발신자 추적이 어려워진다.

취약성은 "대중이 어리석어서" 생기는 것이 아니다. 시간·정보·신뢰가 제한된 환경에서 누구나 출처 휴리스틱과 사회적 신호를 사용한다. 흑색선전은 정상적 인간 판단의 비용절감 메커니즘을 악용한다.

8. 출처 세탁(source laundering)과 디지털 흑색선전

8.1 출처 세탁이란 무엇인가

출처 세탁은 한 메시지가 여러 중간자와 플랫폼을 거치면서 실제 기원과 후원 관계가 흐려지고, 최종 수용자에게는 독립적으로 확인된 이야기처럼 보이는 현상이다. "익명 계정 -> 위장 언론 -> 프런트 조직 -> 가짜 전문가 -> 인플루언서 -> 온라인 커뮤니티 -> 전통 언론 재인용"의 어느 단계에서도 원래 출처 정보가 사라질 수 있다. 이것은 돈세탁처럼 단일 기술이 아니라 provenance(기원 정보)의 연쇄적 손실·변형 문제로 보는 편이 정확하다.

8.2 탐지·방어 관점의 포렌식 포인트

- 최초 게시 시점: 가장 오래된 공개 사본과 최초 유통 계정을 찾고, 이후의 재인용과 구분한다.
- 발신자 실체: 계정·매체·조직이 실제 인물·법인·편집부와 연결되는지 독립 자료로 확인한다.
- 동시성·동일성: 서로 관련 없어 보이는 계정이 매우 짧은 시간 안에 동일 문구·URL·이미지를 반복하는지 본다.
- 원본성: 이미지·영상·문서의 원본 파일, 촬영·작성 맥락, 버전, 메타데이터, 편집 흔적을 확인한다.

- 인프라 연관성: 위장 사이트들이 공통 기술 인프라·도메인 패턴·분석 코드·인증서 흔적을 공유하는지는 귀속의 보조 신호가 될 수 있다.
- 언어 흔적: 번역투·시차·문화적 오류는 단서가 될 수 있으나 단독으로 국적을 확정하는 증거는 아니다.
- 확산 순서: fringe forum -> 가짜 뉴스 -> 소셜 계정 -> 인플루언서 -> 주류 언론이라는 시간적 경로를 재구성한다.
- 독립 확인: 서로 소유·취재망이 다른 신뢰할 만한 매체나 1 차 자료가 같은 사실을 확인하는지 본다.
- 보존: 원본 URL, 화면, 파일, 해시, 타임스탬프를 보존해 삭제 이후에도 검증할 수 있게 한다.
- provenance 표준: C2PA Content Credentials 같은 출처·편집 이력 표준은 방어 수단이지만, 자격정보가 없다는 사실 자체가 곧 가짜라는 뜻은 아니다. [R22][R21]

8.3 현대 디지털 연결구조

소셜미디어 -> 알고리즘 추천 -> 봇·시빌 네트워크 -> 가짜 페르소나 -> AI 생성 텍스트·이미지·음성·영상 -> 딥페이크 -> 마이크로타기팅 -> 인플루언서 -> 온라인 커뮤니티 -> 전통 언론 재인용의 연결은 "조작된 이야기가 자발적 여론처럼 보이는" 조건을 만든다. 그러나 이 사슬을 항상 하나의 중앙집중 작전으로 가정하면 안 된다. 실제 이용자가 자발적으로 재전파하는 순간 작전 메시지와 유기적 정치 행동이 섞이고, 귀속과 효과 측정이 훨씬 어려워진다.

2024 년 미국 DOJ 의 러시아 AI 강화 봇팜 사건은 AI 가 가짜 소셜 페르소나를 대량 생산하는 데 사용될 수 있음을 보여준다. 같은 해 Doppelganger 사건은 위장 언론 도메인, 가짜 시민 프로필, 인플루언서, 유료 광고, AI 생성 광고 일부가 결합된 것으로 법무부가 주장했다. 이는 "생성형 AI 가 선전의 모든 것을 바꿨다"기보다 기존의 사칭·도메인 모방·확산 네트워크를 더 저렴하고 빠르게 만드는 방향으로 작동하고 있음을 보여준다. [R16][R15]

9. 흑색선전과 false flag: 물리적 작전과 커뮤니케이션상의 허위 귀속을 구분하라

false flag 라는 말은 역사적으로 적국·제 3 국의 깃발·표식·정체성을 사용해 행위자를 속이는 군사·해상·정보 활동을 가리켜 왔다. 커뮤니케이션에서 false flag communication 또는 false attribution 은 메시지의 발신자를 다른 집단으로 가장하는 것을 의미한다. 흑색선전은 후자와 상당히 겹치지만, 모든 군사적 false flag operation 이 선전은 아니고 모든 흑색선전이 물리적 작전을 포함하는 것도 아니다.

구분	군사적 false flag operation	커뮤니케이션 false attribution / 흑색선전
대상	적의 전술·군사 판단	대중·군인·언론·정책결정자의 인식
수단	깃발·제복·표식·플랫폼·행동	라디오·문서·사이트·계정·페르소나
목적	기습·접근·오도·위장	신뢰 탈취·분열·여론·행동 변화
법적 핵심	전쟁법상 perfidy·표식 오용 규칙 검토	표현의 자유·사기·선거법·FARA·플랫폼 규칙 등
주의	적 제복 사용이 언제나 perfidy 인 것은 아니며 구체적 사용 맥락이 중요	익명성 자체는 false attribution 이 아님

국제인도법상 Additional Protocol I 제 37 조는 적을 죽이거나 부상시키거나 포획하기 위해 법적 보호를 받을 것이라는 신뢰를 배신하는 perfidy 를 금지하지만, 국제법을 위반하지 않는 camouflage, decoys, mock operations, misinformation 같은 ruses of war 는 금지하지 않는다고 규정한다. 제 39 조는 공격 중 적국의 깃발·군사표식·제복을 사용하는 것을 제한한다. 따라서 "선전에서 출처를 속였다"는 이유만으로 자동으로 perfidy 가 되는 것은 아니다. [R29][R30]

10. 민주주의·선거에서의 흑색선전

후보자 평판 공격

가짜 내부고발자·위장 언론·조작 문서가 후보자의 범죄, 건강, 부패, 외세 연계 등을 주장하면 언론은 진위 확인에 자원을 쓰고 후보는 반박 프레임에 갇힌다.

투표 억제

특정 집단을 겨냥해 투표일·장소·자격·우편투표 절차에 관한 거짓 정보를 퍼뜨리면 설득이 아니라 참여 자체를 줄이는 효과를 노릴 수 있다.

집단 간 불신 증폭

인종·종교·이민·치안·젠더 같은 기존 갈등에 가짜 시민 페르소나를 투입하면 외부 개입이 내부 갈등처럼 보인다.

선거 정당성 훼손
"결과는 이미 조작됐다", "모든 기관이 한통속"이라는 메시지는 특정 후보의 승패보다 장기적 제도 신뢰를 약화할 수 있다.

정치적 냉소주의
상반된 거짓말을 동시에 뿌리는 작전은 특정 믿음을 만들기보다 "진실은 알 수 없다"는 무기력과 disengagement 를 만들 수 있다.

언론·기관 불신
위조문서나 가짜 뉴스사이트가 실제 언론 로고를 사칭하면 발각된 뒤에도 원래 기관의 신뢰가 함께 손상될 수 있다.

미국의 2018 년 IRA 기소문은 러시아 조직이 미국인처럼 보이는 소셜 계정과 조직을 만들고 정치적 활동을 촉진했다고 주장하지만, 그 기소는 2016 년 선거 결과가 변경됐다고 주장하지 않았다. 영향작전의 존재, 노출 규모, 투표 선택의 변화, 선거 결과의 변화는 서로 다른 명제다. [R14]

AI 생성 개념 삽화: 선거·시장에 대한 보이지 않는 영향

11. 기업·금융시장에서의 유사점과 차이
기업 평판 공격, 경쟁사 위장 캠페인, 허위 루머, 가짜 고객·전문가, 위장 연구기관, 조작된 "유출" 자료는 정치적 흑색선전과 구조적으로 유사하다. 핵심은 제 3 자 독립성이나 자발적 여론을 가장해 메시지의 신뢰를 높이는 것이다. 다만 금융시장에서는 정보가 가격·거래·재산 손실과 직접 연결되므로 증권사기·시장조작 규제가 더 즉각적으로 작동한다.

비교	정치적 흑색선전	기업·금융시장형 기만
주요 목표	여론·선거·정책·동맹·사기	평판·구매·투자 판단·가격
핵심 수단	가짜 시민·위장 언론·프런트 조직	가짜 고객·분석가·연구·언론·루머
시간축	장기적 신뢰 침식 가능	분·시간 단위 가격 반응도 가능
법적 위험	선거법·FARA·사기·명예훼손·국가안보	Rule 10b-5, Exchange Act §9, 사기·명예훼손·상표법 등
효과 측정	태도·투표·신뢰의 인과가 어려움	가격·거래량은 측정 쉽지만 인과·의도 입증은 별개

미국 증권법상 Exchange Act §9 는 허위·오도되는 시장 외관과 특정 조작적 행위를 금지하고, Rule 10b-5 는 증권 매매와 관련한 사기 장치, 중요 사실의 허위진술·누락, 사기적 관행을 금지한다. 본 보고서는 그러한 불법적 시장조작을 실행하는 방법을 다루지 않는다. [R34][R35]

12. 미국법·플랫폼 정책·국제법의 쟁점

12.1 명예훼손과 First Amendment
미국에서 허위 발언이라고 해서 자동으로 형사처벌할 수 있는 것은 아니다. United States v. Alvarez 에서 연방대법원은 복수의견과 동의의견을 통해 단순한 거짓말 자체를 포괄적으로 비보호 범주로 만들 수 없다는 결론에 이르렀다. 반면 사기, 위

증, 특정 명예훼손 등 구체적 법익 침해와 결합된 거짓말은 규제될 수 있다. 공직자에 대한 명예훼손은 New York Times v. Sullivan 에 따라 허위성뿐 아니라 actual malice, 즉 거짓임을 알았거나 진실 여부를 무모하게 무시했다는 높은 기준이 적용된다. [R26][R25]

12.2 사기·신원 위조·상표·도메인

위장 사이트·가짜 신분·사칭이 금전·재산 취득을 위한 사기 계획과 결합되면 wire fraud 등 일반 형사법이 적용될 수 있다. 2024 년 Doppelganger 사건에서 DOJ 는 러시아 연계 위장 사이트들이 실제 미국 언론의 도메인과 상표를 모방했다고 주장하며 돈세탁·상표 관련 법률 등으로 도메인 압수를 추진했다. [R15][R36]

12.3 선거법: 외국인의 선거 관련 지출·기여

52 U.S.C. §30121 은 외국인이 연방·주·지방 선거와 관련해 기부·기여, 독립지출, 선거운동 통신 지출 등을 하는 것을 금지한다. 구체적 온라인 발언·서비스가 "기여"나 "thing of value"에 해당하는지는 사실관계와 판례에 따라 복잡해질 수 있으므로 개별 사안은 전문 법률검토가 필요하다. [R27]

12.4 FARA 와 외국 영향력 공개

Foreign Agents Registration Act(FARA)는 외국 principal 의 지시·통제 아래 미국 내에서 일정한 정치활동, 홍보·정보 서비스 등을 하는 agent 에게 등록·관계 공개를 요구한다. DOJ 는 FARA 를 외국 주체와의 관계·활동·재정 흐름을 공개하는 disclosure statute 로 설명한다. 역사적으로 1938 년 법의 출발점은 외국 선전 문제였다. [R28]

12.5 플랫폼 정책

Facebook·Instagram, YouTube, X, TikTok 등 민간 플랫폼은 impersonation, coordinated inauthentic behavior, spam, manipulated media, deceptive practices 등에 자체 규칙을 둔다. 이 규칙은 법률과 동일하지 않고 자주 변경된다. Meta 는 Spamouflage 를 "coordinated inauthentic behavior" 차원에서 제거했고, Doppelganger 도 실제 언론을 모방하는 네트워크로 다뤘다. 플랫폼 집행 결과는 중요한 증거이지만 정부 귀속의 법적 확정과 동일하지 않다. [R18]

12.6 국가안보와 표현의 자유의 긴장

민주국가의 방어는 두 위험 사이 균형을 요구한다. 외국 비밀 영향공작을 방치하면 선거·안보가 침해될 수 있지만, "허위정보"를 넓게 정의해 정부가 국내 정치적 발언을 규제하면 반대 의견·풍자·오류·미확인 주장까지 위축시킬 수 있다. 따라서 규제의 초점은 가능한 한 행위 기반(외국 대리 관계 은폐, 신원 사칭, 사기, 불법 자금, 해킹, 협박, 조작적 네트워크)과 투명성에 두고, 진실 판정만으로 형사화하는 접근은 높은 헌법적 위험을 가진다.

12.7 국제법·전쟁법

전시 선전·기만은 그 자체로 전면 금지되지 않는다. 국제인도법은 합법적 ruse 와 perfidy 를 구분한다. 보호받는 지위나 항복 의사를 가장해 적을 죽이거나 포획하는 행위는 금지되지만, 국제법을 위반하지 않는 misinformation·decoy·camouflage 는 ruse 가 될 수 있다. 또한 적국 제복·표식의 사용은 공격 등 특정 상황에서 금지된다. 사이버·정보공간의 false attribution 이 언제 무력공격·간섭·주권 침해에 해당하는지는 행위의 효과와 수단에 따라 별도 국제법 분석이 필요하다. [R29][R30]

13. 흑색선전 탐지 프레임워크: 10 개 핵심 질문

질문	검증 포인트
1. 최초 출처는 누구인가?	현재 보고 있는 게시물이 아니라 가장 이른 공개 사본·원본 계정·원본 문서를 추적한다.
2. 출처를 독립적으로 확인할 수 있는가?	소개 페이지·자기소개만 보지 말고 법인·기자·전문가 경력·기관 기록 등 외부 자료로 실체를 확인한다.
3. 발신자의 정체성이 실제인가?	프로필 사진, 과거 활동, 네트워크, 오프라인 실체, 다른 플랫폼과의 일관성을 본다.
4. 최초 게시 시점과 확산 경로는 무엇인가?	누가 먼저 올렸고 어떤 계정·사이트·인플루언서·언론이 이어받았는지 타임라인을 만든다.
5. 비정상적으로 동일한 메시지가 반복되는가?	문구·해시태그·URL·이미지·게시시간의 동시성을 본다. 단, 자연스러운 캠페인·보도자료 확산과 구분한다.
6. 감정을 과도하게 자극하는가?	공포·분노·혐오·수치·배신을 즉시 유발하면서 검증할 시간과 맥락을 제거하는지 본다.
7. 증거보다 정체성·분노·공포에 의존하는가?	"우리 편이면 믿어야 한다"는 압력이 강할수록 일단 주장을 분리해 검증한다.
8. 독립적인 정보원이 확인하는가?	같은 보도자료를 베낀 여러 사이트는 독립 확인이 아니다. 취재·자료의 독립성을 본다.
9. 이미지·영상·문서 원본성을 확인할 수 있는가?	원본 파일, 이전 버전, 촬영 위치·시간, 문서 형식, Content Credentials 등 provenance 를 확인한다. [R21][R22]
10. 누가 전략적 이익을 얻는가?	동기 분석은 보조 증거다. 이익을 얻는다는 사실만으로 발신자를 확정하지 않는다.

13.1 실무 체크리스트

- ☐ 원본 URL·파일·스크린샷·게시시각을 보존했다.
- ☐ 최초 출처와 현재 재전파 출처를 구분했다.
- ☐ 발신자·매체·조직의 실체를 외부 자료로 검증했다.
- ☐ 주장(claim), 증거(evidence), 해석(frame), 출처(attribution)를 각각 분리했다.
- ☐ 최소 2 개 이상의 독립적 1 차·신뢰 매체가 핵심 사실을 확인하는지 봤다.
- ☐ 이미지 역추적·영상 프레임·문서 버전·메타데이터 등 원본성을 확인했다.
- ☐ 계정망의 생성 시점·활동 패턴·동시 게시를 확인했다.
- ☐ 번역문이면 원문과 비교했다.
- ☐ 감정적 반응이 강할수록 공유·매매·투표 관련 행동을 미루고 검증했다.
- ☐ 귀속 수준을 "확인", "고신뢰 평가", "가능성", "미확인" 등으로 구분했다.
- ☐ 발각된 작전의 규모와 실제 효과를 구분했다.
- ☐ 반박할 때 원래 거짓 주장을 과도하게 반복해 재확산하지 않도록 했다.

AI 생성 개념 삽화: 출처와 원본성의 포렌식 검증

13.2 방어 원칙

- 빠른 선점(prebunking): 예상되는 조작 유형과 검증법을 사전에 교육

- 출처 투명성: 언론·정부·연구기관이 자료·방법·수정 이력을 공개
- 다중 독립 채널: 단일 기관의 권위보다 서로 독립적인 검증망 구축
- 증거 중심 반박: 공격자의 의도를 추측하기보다 검증 가능한 사실·원본을 먼저 제시
- 비례 대응: 영향이 거의 없는 작전을 과도하게 홍보해 오히려 증폭시키지 않기
- 플랫폼·언론·연구자 협력: 계정·도메인·콘텐츠·타임라인을 공유하되 시민 감시와 투명성 유지
- 미디어 리터러시: "무엇을 믿을까"보다 "어떻게 출처를 확인할까"를 훈련

14. 분석상의 오류와 효과 측정의 한계

오류	교정 원칙
"내가 싫어하는 주장 = 선전" 오류	선전 여부는 정치적 호감이 아니라 목적, 조직성, 메시지 설계, 발신자 관계 등 증거로 판단해야 한다.
"익명 = 정보기관" 오류	익명 계정은 개인·취재원 보호·활동가·사기꾼·마케팅·국가행위자 등 다양한 원인을 가진다. 익명성만으로 귀속 불가.
"거짓 내용 = 흑색선전" 오류	출처가 공개된 거짓말은 disinformation 일 수 있어도 흑색선전은 아닐 수 있다.
"가짜 출처 = 거짓 내용" 오류	흑색선전은 실제 사실을 사용해 더 강한 신뢰를 얻을 수 있다.
"계정 수 = 영향력" 오류	Spamouflage 와 Secondary Infektion 처럼 수천 계정·수백 플랫폼을 사용해도 진짜 참여가 낮을 수 있다. [R17][R18]
"노출 = 설득" 오류	메시지를 봤다는 것, 믿었다는 것, 행동했다는 것, 선거·전쟁 결과가 바뀌었다는 것은 각각 별도 검증이 필요.
"정치적 이익 = 행위자" 오류	누가 이익을 얻는지는 동기 단서일 뿐 기술·인적·문서 증거 없이 귀속을 확정할 수 없다.
"플랫폼 발표 = 사법적 사실 확정" 오류	플랫폼의 threat intelligence 는 중요하지만 증거 기준과 공개 범위가 사법절차와 다르다.
"정부 발표 = 자동으로 확정" 오류	정보기관·법집행 발표도 공개된 증거 수준, 법적 상태(혐의/기소/유죄), 반대 평가를 함께 봐야 한다.
"팩트체크만 하면 해결" 오류	핵심이 출처·정체성 위장이라면 사실 하나를 바로잡아도 네트워크와 신뢰 구조가 남을 수 있다.

15. AI 와 생성형 미디어가 흑색선전을 어떻게 바꿀 것인가

15.1 비용 하락과 규모 확대

생성형 AI 는 다국어 문장, 지역별 변형, 프로필 이미지, 합성 음성·영상 제작 비용을 낮춘다. 특히 "한 사람이 운영하는 수많은 서로 다른 페르소나"를 자연스럽게 유지하는 비용이 감소하면 source falsification 의 규모가 커질 수 있다. 2024 년 DOJ 의 AI 강화 러시아 봇팜 사건은 이러한 방향을 보여주는 초기 사례다. [R16]

15.2 정교한 개인화보다 "충분히 그럴듯한 대량 변형"이 먼저 확산될 가능성

AI 의 가장 즉각적인 위험은 완벽한 딥페이크보다, 보통 수준의 기사·댓글·이미지·음성을 매우 많이 만들어 A/B 식으로 반응이 좋은 것을 선택할 수 있게 하는 생산성 향상이다. 그러나 실제 설득력은 언어 능력보다 신뢰 관계, 지역 문화, 기존 사회갈등, 유통 채널에 좌우된다. Spamouflage 의 낮은 유기적 참여는 "콘텐츠를 많이 만들 수 있다"와 "사람을 설득할 수 있다"가 다르다는 점을 보여준다. [R18]

15.3 딥페이크보다 더 큰 문제: provenance 붕괴와 liar's dividend

합성물이 흔해지면 진짜 영상도 "AI 로 만든 것"이라고 부인하기 쉬워지는 liar's dividend 가 커질 수 있다. 따라서 방어는 단순 딥페이크 탐지 모델에만 의존하기보다 촬영·편집·배포 과정의 provenance 를 인증하는 방향으로 이동해야 한다. NIST 는 합성콘텐츠 위험 완화를 위해 출처 인증, 워터마킹, 탐지, 테스트·감사를 함께 검토하고 있으며 C2PA 는 미디어의 출처와 편집 이력을 검증하는 공개 표준을 개발한다. [R21][R22]

15.4 AI 에이전트와 자동화된 출처 세탁

향후에는 하나의 모델이 글을 쓰는 수준을 넘어, 다수 에이전트가 서로 다른 페르소나를 유지하고 다른 플랫폼의 논쟁에 반응하며 번역·요약·밈 제작을 자동화할 수 있다. 방어 측에서도 AI 가 계정 간 동시성, 텍스트 유사도, 네트워크 이상징후, 합성 미디어 흔적을 탐지하는 데 쓰일 것이다. 공격과 방어가 모두 자동화되면서 중요한 자산은 "콘텐츠 진위 판정"보다 신뢰 가능한 신원·출처·편집 이력 인프라가 될 가능성이 높다.

15.5 전망의 핵심

미래의 흑색선전은 "더 그럴듯한 거짓말"만의 문제가 아니다. 핵심 경쟁은 누가 말했는지, 실제 사람이 존재하는지, 원본이 어디서 왔는지, 여러 출처가 정말 독립적인지를 검증할 수 있는가에 있다. 즉 AI 시대의 중심 문제는 truth crisis 만이 아니라 provenance crisis 다.

부록 A. 흑색선전 핵심 개념 20 개

개념	정의
1. Black propaganda	실제 발신자를 다른 주체로 허위 귀속시키는 선전.
2. Source deception	발신자·후원자·기원을 속이는 행위.
3. Source falsification	출처 정보를 거짓으로 표시·구성하는 것.
4. False attribution	행위·메시지를 실제와 다른 행위자에게 귀속시키는 것.
5. White propaganda	실제 출처가 공개·정확히 식별되는 선전.
6. Grey propaganda	출처가 불명확·비공개·모호한 선전.
7. Disinformation	기만 의도를 가진 허위·오도 정보.
8. Misinformation	기만 의도 없이 공유되는 거짓·부정확 정보.
9. Malinformation	진짜 정보의 해로운 공개·맥락 왜곡.
10. Covert influence	후원·지휘를 숨긴 영향력 활동.
11. Front organization	실제 통제자를 숨긴 명목상 독립 조직.
12. Astroturfing	조직된 활동을 자발적 풀뿌리 운동처럼 보이게 함.
13. Sockpuppet	한 운영자가 독립된 다른 사람인 척 사용하는 온라인 정체성.
14. Sybil network	다수의 가짜·중복 정체성이 독립 행위자처럼 행동하는 네트워크.
15. Source laundering	재전파를 통해 기원·후원 정보가 지워지고 독립적 출처처럼 보이는 과정.
16. Forged document	발신자·서명·형식·내용 등을 위조한 문서.
17. Imposter content	실제 인물·기관·언론의 이름·형식을 사칭하는 콘텐츠.
18. Coordinated inauthentic behavior	실제 정체성·협력 관계를 속이며 조직적으로 행동하는 계정·페이지 네트워크를 가리키는 플랫폼 용어.
19. False flag communication	상대·제 3 자가 보낸 메시지처럼 위장하는 커뮤니케이션.
20. Provenance	콘텐츠가 누가 언제 어떻게 만들고 수정·배포했는지에 관한 기원·이력 정보.

부록 B. 주요 연구자·이론가 10 명

연구자	핵심 기여
Harold D. Lasswell	1 차대전 선전 연구의 고전. 선전의 내용·채널·효과를 체계적으로 분석.
Edward Bernays	대중관계와 여론 형성의 실무·이론을 대중화. 선전·PR 의 윤리 논쟁과 함께 읽어야 함.
Jacques Ellul	현대 대중사회·기술사회에서 선전의 사회학적 구조를 분석.
Paul M. A. Linebarger	Psychological Warfare 의 고전적 군사·정치 커뮤니케이션 이론가.
Garth S. Jowett	Victoria O'Donnell 과 함께 현대 propaganda 연구의 대표 교과서 체계를 구축.
Victoria O'Donnell	선전의 정의, 분석 모델, 역사 사례를 체계화한 공동 저자.
Philip M. Taylor	전쟁·외교·미디어와 선전의 장기 역사를 연구한 대표 학자.
Thomas Rid	Active Measures 에서 냉전부터 디지털 시대까지 비밀 정치전·허위정보 역사를 분석.
Claire Wardle	misinformation/disinformation/malinformation 을 구분한 information disorder 프레임워크의 공동 저자.

연구자	핵심 기여
Ben Nimmo	현대 OSINT 기반 영향작전·가짜 계정·cross-platform attribution 연구를 발전시킨 조사자.

부록 C. 권위 있는 학술서·보고서·정부자료 20+

자료	활용 가치
[R1] Garth S. Jowett & Victoria O'Donnell, Propaganda & Persuasion, SAGE, 최신판.	선전의 정의·역사·분석 모델의 표준 교과서.
[R2] Jacques Ellul, Propaganda: The Formation of Men's Attitudes, Vintage.	대중사회와 선전의 구조적 분석.
[R3] Paul M. A. Linebarger, Psychological Warfare, 1948/1954 editions.	심리전과 흑색·회색 선전의 고전적 실무·이론.
[R4] Philip M. Taylor, Munitions of the Mind: A History of Propaganda, Manchester University Press.	전쟁과 선전의 장기 역사.
[R5] Sefton Delmer, Black Boomerang, 1962.	PWE 흑색 라디오 실무자의 회고록. 1 차 사료 성격과 자기서술의 한계를 함께 고려.
[R6] Ellic Howe, The Black Game: British Subversive Operations Against the Germans During the Second World War, 1982.	영국 흑색선전 작전의 역사.
[R7] CIA, "The Spymaster's Toolkit." https://www.cia.gov/stories/story/the-spymasters-toolkit/	CIA 가 white/grey/black propaganda 를 설명하는 공개 역사자료.
[R8] CIA, "Barbara Lauwers: Deceiving the Enemy." https://www.cia.gov/stories/story/barbara-lauwers-deceiving-the-enemy/	OSS Morale Operations 와 Operation Sauerkraut 사례.
[R9] CIA, "Spy Girl Betty McIntosh." https://www.cia.gov/stories/story/spy-girl-betty-mcintosh/	OSS 의 rumor·false news·radio black propaganda 사례.
[R10] U.S. Department of State, FRUS 1969-1976, Vol. VII, Document 77. https://history.state.gov/historicaldocuments/frus1969-76v07/d77	CIA 의 베트남 black radio/letters/leaflets 를 명시한 기밀해제 문서.
[R11] U.S. Department of State, FRUS 1969-1976, Vol. VIII, Document 144. https://history.state.gov/historicaldocuments/frus1969-76v08/d144	1972 년 covert psychological warfare 보고.
[R12] Claire Wardle & Hossein Derakhshan, Information Disorder, Council of Europe, 2017. https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html	mis/dis/mal-information 프레임워크.
[R13] Christopher Paul & Miriam Matthews, The Russian "Firehose of Falsehood" Propaganda Model, RAND, 2016. https://www.rand.org/pubs/perspectives/PE198.html	고용량·다채널·반복·비일관성 모델과 심리 연구.
[R14] U.S. DOJ, Grand Jury Indicts Thirteen Russian Individuals and Three Russian Companies, 2018. https://www.justice.gov/archives/opa/pr/grand-jury-indicts-thirteen-russian-individuals-and-three-russian-companies-scheme-interfere	IRA 의 미국인 사칭·소셜미디어 영향활동에 대한 형사사건 자료.
[R15] U.S. DOJ, "Doppelganger" domain seizures, Sept. 4, 2024. https://www.justice.gov/archives/opa/pr/justice-department-disrupts-covert-russian-government-sponsored-foreign-malign-influence	위장 언론 도메인·가짜 미국인 계정·인플루언서·AI 광고 사례.
[R16] U.S. DOJ, AI-Enhanced Russian Social Media Bot Farm, July 9, 2024. https://www.justice.gov/archives/opa/pr/justice-department-leads-efforts-among-federal-international-and-private-sector-partners	AI 를 이용한 가짜 소셜 페르소나 네트워크 사례.
[R17] Graphika, Exposing Secondary Infektion, 2020. https://graphika.com/reports/exposing-secondary-infektion	위조문서·가짜 계정·300+ 플랫폼, 낮은 실제 영향의 중요한 비교 사례.
[R18] Meta, "Raising Online Defenses Through Transparency and Collaboration," Aug. 2023. https://about.fb.com/news/2023/08/raising-online-defenses/	Spamouflage 와 Doppelganger 관련 플랫폼 threat intelligence.
[R19] Global Affairs Canada, Spamouflage campaign assessment. https://international.canada.ca/en/global-affairs/corporate/reports/rapid-response-mechanism/news/2023-china-spamouflage	AI 생성 프로필·조작 콘텐츠와 중국 연계 평가.
[R20] European External Action Service, 4th Annual Report on FIMI Threats, 2026. https://www.eeas.europa.eu/eeas/4th-eeas-annual-report-foreign-information-manipulation-and-interference-threats_en	최신 EU FIMI 위협 프레임워크.
[R21] NIST AI 100-4, Reducing Risks Posed by Synthetic Content, 2024. https://www.nist.gov/publications/reducing-risks-posed-synthetic-content-overview-technical-approaches-digital-content	provenance, watermarking, detection 등 합성콘텐츠 방어기술 검토.
[R22] C2PA Specifications 2.4. https://spec.c2pa.org/specifications/	디지털 콘텐츠 출처·편집 이력 인증 표준.
[R23] S. Vosoughi, D. Roy, S. Aral, "The spread of true and false news online," Science 359 (2018): 1146-1151. doi:10.1126/science.aap9559	거짓 뉴스 확산의 속도·범위와 인간 공유 행동 연구.
[R24] G. Pennycook, T. Cannon, D. Rand, "Prior exposure increases perceived accuracy of fake news," JEP: General 147 (2018): 1865-1880. doi:10.1037/xge0000465	반복 노출과 진실착각효과.
[R25] New York Times Co. v. Sullivan, 376 U.S. 254 (1964). https://www.law.cornell.edu/supremecourt/text/376/254	공직자 명예훼손의 actual malice 기준.
[R26] United States v. Alvarez, 567 U.S. 709 (2012).	거짓 발언 자체에 대한 포괄적 형사규제와 First Amendment 의 긴장.
[R27] 52 U.S.C. §30121. https://www.law.cornell.edu/uscode/text/52/30121	외국인의 선거 관련 기부·지출 규제.

자료	활용 가치
[R28] U.S. DOJ, Foreign Agents Registration Act. https://www.justice.gov/nsd-fara	외국 principal 과 agent 의 정치활동·정보활동 공개 체계.
[R29] ICRC, Additional Protocol I, Article 37 - Prohibition of perfidy. https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-37	perfidy 와 ruses of war 구분.
[R30] ICRC, Additional Protocol I, Article 39 - Emblems of nationality. https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-39	적국·중립국 표식·제복 사용 제한.
[R31] Milton Leitenberg, "China's False Allegations of the Use of Biological Weapons by the United States during the Korean War," Wilson Center CWIHP Working Paper 78, 2016.	한국전쟁 생물전 주장 조작설의 주요 연구.
[R32] Stephen Endicott & Edward Hagerman, "The Soviets and allegations of biological warfare in the Korean War," 1998 response. https://www.yorku.ca/sendicot/12SovietDocuments.htm	Leitenberg/Weathersby 자료의 출처·해석에 대한 반론.
[R33] Milton Leitenberg, "Resolution of the Korean War biological warfare allegations," Critical Reviews in Microbiology 24(3), 1998. doi:10.1080/10408419891294271	한국전쟁 생물전 주장의 반증을 주장하는 학술 논문.
[R34] 15 U.S.C. §78i - Manipulation of security prices. https://www.law.cornell.edu/uscode/text/15/78i	시장조작·허위 외관 관련 증권법 조항.
[R35] 17 C.F.R. §240.10b-5. https://www.law.cornell.edu/cfr/text/17/240.10b-5	증권 관련 사기·중요 허위진술·기만행위 금지.
[R36] 18 U.S.C. §1343 - Wire fraud. https://www.law.cornell.edu/uscode/text/18/1343	전자통신을 이용한 사기 관련 일반 형사법.
[R37] Thomas Rid, Active Measures: The Secret History of Disinformation and Political Warfare, Farrar, Straus and Giroux, 2020.	20 세기·디지털 시대 비밀 정치전과 disinformation 역사.
[R38] Yochai Benkler, Robert Faris, Hal Roberts, Network Propaganda, Oxford University Press, 2018.	미국 미디어 생태계·네트워크 구조와 정치 정보 문제.
[R39] Peter Pomerantsev, This Is Not Propaganda, PublicAffairs, 2019; How to Win an Information War, 2024.	현대 정보전과 Sefton Delmer 의 역사적 비교.
[R40] U.S. Department of State, Global Engagement Center, How the People's Republic of China Seeks to Reshape the Global Information Environment, 2023.	중국의 국제 정보환경 영향·선전·검열·disinformation 에 대한 미국 정부 평가; 중국 측은 이러한 귀속을 반박해 왔음.

부록 D. 비교표 3 종

D-1. 백색·회색·흑색선전 비교표

항목	백색	회색	흑색
출처	정확히 공개	불명·모호·비공개	의도적으로 다른 주체로 위장
내용 진위	보통 사실 중심이나 선택·프레임 가능	참·거짓·혼합	참·거짓·혼합
기만	낮음	중간	출처 기만이 핵심
신뢰 전략	공식 신뢰·평판	모호성·책임 회피	상대·제 3 자의 신뢰를 탈취
발각 위험	내용 오류 중심	후원관계 노출	사칭·위조·네트워크 귀속 노출
대표 방어	팩트체크·정책 검증	출처 투명성	출처 포렌식·네트워크 분석

D-2. disinformation 과 black propaganda 비교표

항목	Disinformation	Black propaganda
핵심 질문	정보가 의도적으로 거짓·오도인가?	누가 보낸 것처럼 보이게 했는가?
분류 축	내용 + 기만 의도	출처 + 귀속 기만
진짜 정보 사용	좁은 정의에서는 제외될 수 있음	가능
진짜 출처 공개	가능	정의상 핵심에서 벗어남
가짜 페르소나	필수 아님	전형적 수단
관계	흑색선전과 상당히 중첩	disinformation 을 포함할 수도, 사실 기반일 수도 있음

D-3. 전통적 흑색선전과 AI·SNS 시대 흑색선전 비교표

항목	전통적 시대	AI·SNS 시대
매체	라디오·전단·우편·위장 신문	SNS·메신저·영상·팟캐스트·위장 사이트
가짜 정체성 비용	높음: 배우·방송시설·인쇄·현지 요원	낮아짐: 대량 프로필·합성 음성·텍스트 자동화
확산 속도	시간·일 단위	초·분 단위
지역화	현지 언어·문화 전문가 필요	AI 가 초안을 돕지만 문화적 오류는 여전히 약점
사회적 증거	라디오 청취자·소문·현장 네트워크	좋아요·공유·댓글·팔로워·봇·시빌
출처 세탁	신문·제 3 국 언론·프런트 조직	cross-platform repost, 인플루언서, 검색·추천, 주류언론 재인용
포렌식	인쇄·방송 송신원·문서	계정·도메인·메타데이터·네트워크·콘텐츠 provenance
주요 방어	방송 차단·반박·대항 선전	OSINT, CIB 탐지, provenance, 신원검증, 미디어 리터러시
한계	비용·도달 범위	대량 생산은 쉬워도 진짜 신뢰·유기적 참여는 여전히 어려움

결론

흑색선전을 이해하는 가장 정확한 문장은 "거짓 정보를 퍼뜨리는 기술"이 아니라 "출처·정체성·신뢰를 조작하여 상대방이 메시지를 다르게 해석하고 행동하도록 만드는 조직적 정보전 메커니즘"이다. 핵심 질문은 언제나 내용과 출처를 분리하는 데서 시작한다: 이 주장은 사실인가, 그리고 이 말을 한 사람은 정말 그 사람인가?

역사적으로 비밀 라디오와 위조 전단이 맡았던 역할을 오늘날에는 위장 뉴스사이트, 가짜 시민 계정, 봇·시빌 네트워크, 인플루언서, 생성형 AI 가 나눠 맡는다. 그러나 기술이 바뀌어도 성공 조건은 놀랍도록 일관적이다. 목표집단의 실제 불만과 문화에 접속하고, 신뢰할 만한 세부 사실을 제공하고, 출처가 "우리 편 내부자" 또는 "독립 제 3 자"라고 믿게 만들어야 한다. 반대로 문화적 어색함, 낮은 유기적 참여, 과도한 반복, 원본성 결여는 작전을 약화한다.

따라서 민주사회와 언론의 방어는 "모든 거짓을 삭제"하는 불가능한 목표보다 출처 투명성, 독립적 확인, 원본 보존, 네트워크 포렌식, 신속한 정정, 법적 절차, 시민의 검증 역량을 강화하는 방향이 더 견고하다. 특히 AI 시대에는 콘텐츠 자체만 보고 진위를 가리는 방식에서 벗어나, 누가 만들었고 어떻게 편집·배포됐는지를 확인하는 provenance 기반 신뢰 인프라가 핵심 공공재가 될 것이다.