3 미·중, AI 안전 대화 채널 '인시던트 라인' 신설 합의

사고 통보 채널 '인시던트 라인'으로 위험 공동 관리에 합의

● 美·中, 뉴욕 고위급 회담에서 'AI 대화' 공식 출범 합의

- 美 베선트 재무장관과 中 허리펑 부총리가 뉴욕 고위급 회담에서 '미·중 AI 대화' 공식화에 합의, 세계 AI 개발을 양분하는 두 나라가 AI 위험을 정기 논의하는 첫 공식 채널이 출범 ('26.9.20.)
- AI 안전사고가 발생하면 양국이 서로 즉시 알릴 수 있는 전용 소통 채널인 이른바 '인시던트 라인'을 개설하는 데에도 합의, '26.11월 中 선전에서 후속 회담을 개최할 예정
- 기술 패권을 다투는 양국이 위험 관리 영역에 한해 협력 채널을 마련한 것으로, 경쟁과 공조가 병존하는 '경쟁 속 안전 공조' 구도를 보여주는 합의

● AI 보안 사고 현실화 속 정상 차원 약속으로 격상

- '26.7월 오픈AI의 AI 모델이 보안 평가 도중 격리된 시험 환경을 벗어나 허깅페이스 시스템에 무단 접근한 사고가 양국을 대화로 이끈 직접적 배경으로 거론
- 시진핑 주석의 국빈 방미 기간('26.9.23~25.) 양 정상이 도출한 8개 성과·양해 항목에 'AI 대화 설립'과 'AI 사고 양자 소통 채널 구축'이 공식 명시, 실무 합의가 정상 차원 약속으로 격상
- AI가 미·중 정상 외교의 의제로 진입하면서, 장관급 실무 합의에 정상 간 공동 발표라는 정치적 무게가 더해져 후속 논의의 추진 기반이 마련

사고 통보·위험 정의 합의와 美 개발사 책임 원칙

(1) AI판 핫라인 '인시던트 라인'으로 사고 공동 관리 첫발

● 정례 대화와 상시 통보 채널의 이중 구조로 설계

- 인시던트 라인은 정례 협의체 '미·중 AI 대화'와 별도로 상시 가동되는 사고 통보 전용 채널로, 정례 대화가 위험 인식을 조율하는 협의 무대라면 인시던트 라인은 긴급 사건 통보 창구
- AI 사고 정보가 국경을 넘어 양국 당국 간에 직접 오가는 상시 장치가 마련된 것으로, AI 사고를 개별국 내부 문제가 아닌 양국 공동 관리 대상으로 본다는 인식을 반영한 결과로 해석

○ AI가 군사와 나란히 미·중 위기관리 정식 축으로 편입

- 양국은 AI 채널과 별도로 군 당국 간 위기 소통·예방 양해각서(MOU)도 조속히 체결하기로 합의, 갈등이 사고로 번지기 쉬운 군사와 AI 두 분야의 위기관리 장치가 나란히 추진되는 양상
- 베선트 장관은 트럼프 대통령과 시진핑 주석의 개인적 관계를 실무 차원으로 연결하려 한다고 언급, 정상 간 신뢰에 의존하던 위기 소통이 군사·AI 두 분야의 상설 채널로 구조화되는 흐름
- 우발적 충돌을 막기 위한 소통 장치가 군사에서 AI 영역으로 확장된 것으로, 갈등 가능성이 큰 분야일수록 대화 창구부터 마련해 온 전통적 위기관리 방식이 AI 분야에 그대로 적용된 사례

(2) 美·中이 함께 정하는 'AI 위험 정의' 프로토콜 논의 개시

○ 위험의 공통 정의 합의가 프로토콜 논의의 출발점

- 베선트 장관은 프로토콜 논의를 시작해 양측이 주요 AI 위험이 무엇인지 합의하는 것이 목표라고 언급, 대응 절차를 정하기에 앞서 위험의 정의부터 맞추는 단계적 접근으로 풀이
- 우선 거론된 위험 유형은 통제 불가능한 AI 에이전트와 사이버 공간의 비국가 행위자로, 상대국의 기술 개발이 아니라 통제를 벗어난 AI와 제3자 악용이라는 공통 위협에 초점을 맞춘 구성
- 규제 체계가 다른 양국이 위험을 부르는 공통 언어부터 만드는 방식으로, 정의 합의가 이뤄지면 인시던트 라인의 통보 대상과 대응 절차 등 후속 규범을 정하는 기준점 역할을 할 것으로 전망

○ '26.11월 선전 후속회담에서 논의 본격화 예정

- 후속회담 시기에는 中 주최 APEC 정상회의가 같은 선전에서 개최되어 트럼프 대통령도 각료를 대동해 참석할 것으로 예상되는 등 AI 논의가 정상 외교 일정과 맞물려 함께 진행될 전망
- AI 개발을 주도하는 G2가 위험의 정의에 먼저 합의할 경우, 선전 회담 결과는 향후 다자 차원에서 진행될 AI 안전 논의에서 위험을 분류하고 대응하는 사실상의 출발점이 될 가능성

(3) 美, 사전 규제 대신 개발사 책임 원칙으로 AI 안전 관리 기조 제시

○ '사고 책임은 개발사'라는 원칙을 공식화

- 베선트 장관은 회담 직후 인터뷰에서 허깅페이스 해킹 사건을 AI 에이전트가 아니라 오픈AI 경영진의 책임이라고 규정, AI가 일으킨 사고라도 책임은 개발 주체에 있다는 원칙을 공식 표명

- 베선트 장관은 기업이 원한다면 개발 속도를 늦출 수 있다며 속도 조절 권한이 기업에 있음을 책임의 근거로 제시, AI 기업의 면책 요구도 국민에게는 나쁜 거래라며 거부 방침을 명확화
- 관련 법 정비 전에 행정부 발언으로 책임 소재의 기준선이 먼저 제시된 것으로, 사고 발생 시 'AI 탓'이라는 방어 논리를 인정하지 않겠다는 美 행정부의 입장이 분명하게 드러난 대목

○ 성장 우선 기조 속 사전 규제 대신 사후 법 집행으로 통제

- 트럼프 대통령은 AI를 산업혁명이나 인터넷보다 큰 성장 동력으로 규정하며 성장을 억누르지 않겠다고 언급, 법 집행 기관이 필요할 때 통제하는 사후 대응 방식을 규제의 기본 틀로 제시
- 미국 내에서는 AI가 인류를 위협할 수 있다며 강력한 규제를 요구하는 진영과 초지능(SI) 개발까지 장려하는 트럼프 대통령의 성장 기조가 맞서며 규제 수위 논쟁이 지속되는 상황
- 사전 규제 중심의 EU 방식과 달리 △성장 촉진 △기업 자기 책임 △사후 법 집행을 조합한 접근이 美 행정부의 AI 거버넌스 방향으로 제시된 것으로, 안전 관리를 기업 자율에 맡기는 구조

AI 공조 첫발에도 곳곳에 남은 한계와 변수

○ 통상·제재 갈등에 연동되는 선택적 공조라는 한계 존재

- AI 협력 합의에도 통상·제재 등 다른 의제에서는 불확실성과 이견이 그대로 남아, 이번 협력이 관계 전반의 화해가 아니라 사안별로 범위를 정한 선택적 공조라는 성격 확인
- 미·중 관세 휴전은 이번 회담에서 논의됐으나 만료('26.11.10.)를 앞두고 연장 합의 여부가 명확히 밝혀지지 않았고, 이란 제재를 두고도 양국이 공개적으로 입장 차를 드러내는 상황
- 선전 회담 시기가 관세 휴전 연장 협상과 겹쳐 AI 대화도 통상 협상 흐름에 따라 진전되거나 지연될 수 있는 구조로, 채널의 지속성이 양국 관계 전반의 안정에 좌우된다는 한계 내포

○ 강제 장치 없는 소통 절차 수준에 머문 초기 합의

- 이번 합의는 양국 간 대화 정례화와 사고 통보 채널 개설이라는 소통 절차 수준에 머물러 있어, 위험 판정 기준이나 이행을 담보할 검증·강제 장치는 아직 마련되지 않은 초기 단계로 평가

- 합의가 선언 단계를 넘어 구속력 있는 국제 규범으로 발전할지는 아직 미정으로, 당분간은 양국 간 이견 속에서도 채널을 계속 가동하며 상호 신뢰를 단계적으로 쌓아 가는 단계에 머물 전망

출처 : Reuters 외(2026.9.)
https://www.reuters.com/world/asia-pacific/us-china-meet-again-ai-safety-two-months-shenzhen-bessent-says-2026-09-21/
https://whytimes.kr/news/view.php?idx=28172
https://www.washingtontimes.com/news/2026/sep/21/us-china-open-communication-line-alarming-ai-incidents/
https://www.fmprc.gov.cn/eng/xw/zyxw/202609/t20260926_12031663.htm
https://openai.com/index/hugging-face-incident-and-the-road-ahead/