

알기쉬운 데이터 분쟁조정 제도 안내



발간사

안녕하십니까? 데이터분쟁조정위원회 위원장 김후곤입니다.

데이터는 디지털 시대에 인공지능의 진화를 촉발하고 산업혁신을 견인하며 국민의 일상을 새롭게 변화시키는 핵심 동인입니다. 최근 챗GPT 열풍 등 초거대 AI를 둘러싼 글로벌 패권 경쟁 속에서 우리나라의 데이터 시장은 지속적으로 성장하고 있습니다.

데이터 산업 시장 규모는 최근 3개년 연평균('20~'22) 11.9% 성장하여 2022년 25조 원을 달성하였으며, 데이터 인력 또한 최근 3개년('20~'22) 연평균 5.4% 증가하여 2022년에는 40만 명을 돌파하였습니다.

데이터 활용이 고도화되고 보편화되는 과정에서 관련 갈등 또한 복잡하고 다양해지고 있습니다. 또한 산업 발전을 위해서 데이터에 대한 부정사용 방지와 보호, 공정하고 합리적인 데이터 활용의 중요성이 지속 대두되고 있습니다.

사인(私人) 간의 데이터의 생산·거래·활용과 관련된 분쟁 전반을 조정하고 피해를 신속히 구제하여 국민의 데이터 권리를 보호하는 데, 데이터분쟁조정위원회가 앞장서겠습니다.

데이터분쟁조정위원회가 분쟁조정 사례를 지속 창출하며 우리나라의 대표적인 데이터 분쟁 조정기관의 지위를 공고히 하고, 데이터 시장 활성화 및 이용 환경의 발전을 위하여 끊임없이 노력해 나가겠습니다.

2023년 12월

데이터분쟁조정위원회 위원장 김 후 곤

목차

I . 데이터 분쟁조정	5
01. 데이터분쟁조정위원회 소개	5
02.데이터 분쟁조정 제도 소개	6
II . 데이터 분쟁조정 유형별 사례	8
01. 데이터 생산	8
02. 데이터 거래	20
03. 데이터 활용	24
[부록]	
01. 데이터 분쟁조정 Q&A	29

I. 데이터 분쟁조정

1. 데이터분쟁조정위원회 소개
2. 데이터 분쟁조정 제도 소개

01 데이터 분쟁조정 위원회 소개

가. 추진배경 및 목적

데이터·AI 산업이 성장하여 데이터 시장규모는 2020년 20조 원에서 2022년 25.1조 원으로, AI 기업수와 매출액은 2020년 933개, 1.9조 원에서 2022년 1,915개, 3.9조 원으로 크게 증가하였습니다. 전 산업 분야에서 핵심 자원인 데이터의 활용이 보편화·고도화되면서 데이터 관련 갈등 또한 복잡·다양해지고 있습니다. 이에 데이터에 대한 부정사용을 방지하고 데이터를 보호하며, 공정·합리적인 데이터 활용 문화를 정착하는 것이 중요하게 되었습니다.

「데이터 산업진흥 및 이용촉진에 관한 기본법」 제34조에 의거하여 데이터분쟁조정위원회가 2023년 10월 12일 출범하였습니다. 데이터 활용 전반에서 발생하는 분쟁을 신속하게 조정하고 피해를 구제하는 조정은 민사소송법에 따른 재판상 화해와 동일한 효력이 발생합니다. 데이터분쟁조정위원회의 적극적 운영을 통해 공정한 데이터 활용 환경 조성, 데이터 이용자의 권리 보호 등에 이바지할 것입니다.

나. 데이터분쟁조정위원회 구성 및 운영

데이터분쟁조정위원회는 데이터의 생산, 거래 및 활용에 대한 학식과 경험이 풍부한 학계, 법조계, 공공, 산업계 등 다양한 분야의 전문가 28인을 조정위원으로 위촉하여 구성하였습니다.

<데이터분쟁조정위원회 위원 현황>

학계	법조계	공공	산업계
10명	11명	3명	4명

조정업무를 효율적으로 수행하기 위하여 위원장은 3인 이내로 조정부를 구성하며, 조정 신청에 대해 사실조사 진행, 조정 전 합의, 조정안 작성 등의 절차를 통해 분쟁해결을 지원하고 있습니다.

02 데이터 분쟁조정 제도 소개

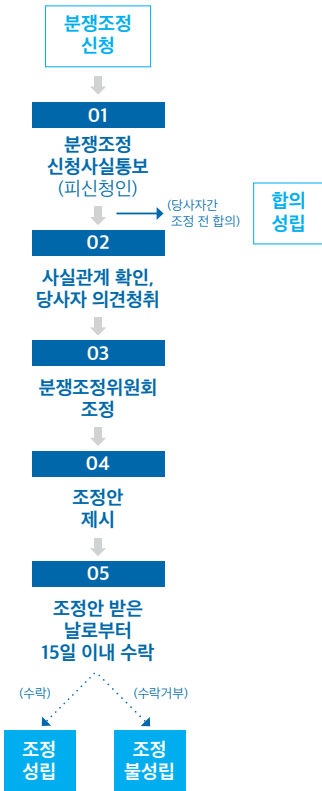
가. 데이터 분쟁조정 기능

데이터 분쟁조정은 소송을 통한 분쟁해결의 단점을 보완하는 대체적 분쟁해결 제도(ADR:Alternative Dispute Resolution)입니다. 보통 데이터와 관련된 갈등이 발생한 경우 다음과 같은 여러 부담 때문에 소송을 포기하고 체념하게 됩니다.

- 소송을 어떻게 시작해야 할지 막막함
- 앞으로 계속 거래를 해야 하는 상대방과의 관계가 어색해짐
- 판결 선례가 없어 승소 예측이 불가능
- 소송에 들이는 노력과 비용에 비해 다투는 금액이 적음
- 소송 기간 예측이 불가능
- 소송에 승소하더라도 시간이 경과하여 실질적인 이익이 적음

데이터 분쟁조정은 간단하게 신청할 수 있고, 제3의 전문가 앞에서 본인과 상대방의 이야기를 나누며 건설적인 타협안을 찾을 수 있고, 데이터 분야 전문가가 제시하는 새로운 해법을 확인할 수 있으며, 짧은 기간 내에 절차가 마무리되어 부담이 크지 않습니다.

< 데이터 분쟁조정 절차도 >



※ 데이터분쟁조정위원회 홈페이지(ddrc.go.kr)에서 신청

나. 데이터 분쟁조정 절차 및 방법

데이터 분쟁의 당사자 누구나 신청할 수 있으며, 데이터분쟁조정위원회는 당사자가 접수한 신청서를 확인하고, 상대방(피신청인)에게 신청 사실을 알려며 답변서 제출을 요청하게 됩니다.

분쟁조정이 접수되면 위원장은 조정위원 3인 이내로 조정부를 구성하고, 조정부는 당사자 자료 요청, 당사자 의견 진술 등을 거쳐 분쟁조정 신청을 받은 날로부터 45일 이내 조정안을 작성하여 당사자에게 통보하게 됩니다. (부득이한 사정이 있는 경우 연장 가능)

양 당사자가 조정안을 수락하는 경우 조정조서 작성과 함께 조정은 성립되고, 수락 여부를 알리지 않거나 거부 시 조정은 불성립됩니다.

II. 데이터 분쟁조정 유형별 사례

01. 데이터 생산

02. 데이터 거래

03. 데이터 활용

01 데이터 생산

쟁점 1.

데이터 생성 시 계약서에 명시되지 않은 전문성 충족 여부



A사와 B사는 3D 아바타 안무가 개발을 위해 필요한 안무 동영상 데이터 공급 계약을 맺었다. B사가 안무 동작을 동영상으로 촬영한 데이터를 A사에 제공하기로 했다. 하지만 자격증을 소지한 안무가 외에 비전문가인 학생들이 생성한 데이터에 대해 A사가 전문성 문제를 제기했고, B사는 '자격증 소지자'가 아닌 '능통자'로 명시된 계약서를 제시해 분쟁이 발생했다.

A사 주장

업체에서 제공받은 데이터의 전문성이 떨어집니다. 계약할 때 데이터 수집 대상을 명확히 하지 않았더니 데이터 제공업체에선 문제가 없다고 하네요….

A사는 K팝 댄스 페스티벌에서 사용할 3D 아바타 안무가를 개발하는 프로젝트에 참여하기로 했다. 세계 각국에서 온 페스티벌 참가자들이 K팝 댄스를 배워 볼 수 있도록, 안무가로부터 동작을 미리 학습한 3D 아바타가 전문 강사를 대신해 영상으로 안무를 시연하는 프로그램을 위한 것이었다. 전반적인 연구개발 분야는 한 대학 연구팀과 협업하기로 하고 A사는 안무 동작 데이터의 수집 및 가공을 담당하기로 했다.

하지만 여러 분야의 데이터를 다뤄온 데이터 전문기업인 A사도 안무와 관련해서는 노하우나 안무전문가 후보군(Pool)이 없었기에 대행업체를 통해 B사를 소개받았다. 그리고 B사와 100개의 안무 동작을 안무 전문가 20명이 각각 녹화해 총 1만 개의 원본 데이터와 3D 아바타 적용을 위한 3D 포인트 데이

터를 제공하는 조건으로 계약을 맺었다.

그런데 3개월 후 최종 결과물을 받아 확인하는 과정에서 의문이 생겼다. 안무 촬영에 참여한 사람들이 전문가로 보이지 않았기 때문이다. 적정성 검토를 요청했더니 B사는 5명의 방송댄스 자격증 보유자와 15명의 댄스 동아리 학생이 촬영을 진행했다고 답변했다.

A사는 댄스 동아리 학생을 '전문가'로 보기 어렵다고 판단하고 B사에 추가 촬영을 요구했다. 그러나 B사는 안무 전문가를 예산 내에서 최대한 섭외했고, 방송댄스 자격증을 보유한 전문가의 감독 하에 촬영을 진행해 정확한 데이터를 전달했다고 반박하며 재촬영을 거부했다.

B사 주장

의뢰받은 대로 데이터를 수집해 제공했습니다. 그런데 계약서상 명시되지 않은 자격증 보유자의 데이터가 아니라며 무상으로 추가 작업을 요구하네요….

B사는 A사로부터 안무 동영상 데이터 수집을 의뢰받았다. A사는 K팝 안무에 능통한 15~20명이 총 100개의 동작을 각각 녹화한 데이터가 필요하며 최대한 많은 데이터를 수집하고 싶다고 했다. 이에 B사는 최종 20명이 1만 건의 안무 동작 데이터를 제공하는 조건으로 A사와 계약을 맺었다.

이후 B사는 곧바로 K팝 안무에 능통한 방송댄스 자격증 보유자와 댄스 동아리 학생 등 20명을 섭외해 촬영을 진행했다. 계약 한 달 정도 되었을 때는 중간 데이터를 A사에 보냈고, 3개월이 지난 시점에 마침내 최종 데이터를 전달한 후 프로젝트 종료를 기다렸다.

그런데 최종 데이터를 확인한 A사가 안무 촬영 참여자의 자질에 대해 이의를 제기해 왔다. A사는 방송댄스 자격증 등을 보유한 안무 전문가가 영상을 녹화하기로 했는데, 비전문가인 학생들도 참여한 것은 계약 위반이라고 주장하며 추가 작업을 요구했다.

하지만 계약서에는 'K팝 댄스 능통자'로 명시되어 있을 뿐 '자격증 보유'에 대한 내용은 없었다. 이를 토대로 B사는 계약 위반이 아니라고 설명하고, 방송댄스 자격증 보유자의 관리 감독 하에 촬영을 진행했다고 설명했다. 또한 오류의 경우 이미 재촬영을 진행해 최종 데이터 전달 전 모두 수정했으며, 촬영 전 가이드 영상도 제작해 참여자들을 숙지시킨 후 작업했음을 부연하고 추가 촬영을 거부했다.



알아두기

데이터 전문성 관련 분쟁

이 사례와 같이 수집된 데이터의 전문성에 관한 분쟁을 피하기 위해서는 계약 시 데이터 품질을 좌우하는 참여 인력의 경력 사항(ex. 자격증 소지 등)을 협의하여 계약서에 명확히 기재할 필요가 있습니다.

01 데이터 생산

쟁점 2.

데이터 가공 중 변수 발생 시 최종 결과물의 완성도 기준



C사와 D사는 여행지 이미지 데이터 라벨링 계약을 맺었다. C사는 업계 관례를 토대로 이미지가 아닌 인스턴스 라벨링 수량을 계약 기준으로 제시했고 D사도 이에 동의했다. 작업 과정에서 D사가 사전에 제시한 평균 인스턴스 수량에 차이가 났지만 C사는 계약한 인스턴스 수량만을 납품했고, D사는 이미지 수량을 기준으로 추가 작업을 요구해 분쟁이 발생했다.

C사 주장

데이터 가공 중 계약 전 의뢰 내용과 다른 점이 발견되어 최종 결과물 수량은 계약서를 기준으로 전달했습니다. 그런데 의뢰자가 무상으로 추가 작업을 요구하네요….

데이터 가공 기업인 C사는 여행 관련 VR 기술을 개발하고 있는 D사로부터 데이터에 대한 Bbox 라벨링 의뢰를 받고 계약을 체결했다. 일반적으로 Bbox 라벨링 작업 및 검수는 1개의 Bbox를 생성하고 검수하는 데 개별적인 시간이 소요되는 점을 감안해 인스턴스 수를 기준으로 계약을 체결한다. 이에 C사는 D사가 4만 장의 이미지 데이터를 확보했으며 평균 인스턴스 수량은 10개라고 한 내용을 토대로 총 40만 개의 인스턴스를 라벨링하기로 합의했다.

그런데 실제 이미지를 받아 작업을 진행해 보니 각 이미지의 평균 인스턴스 수량은 10개가 넘는 10.5개 수준이었다. 즉 4만 건이 아닌 3만 8,096장의 이미지 데이터를 가공해야 계약한 총 40만 개의 인스턴스를 라벨링 할 수 있었다. 이를 기준으로 C사는 3만 8,100장의 이미지 데이터를 라벨링 해 검수까지

마친 후 D사에 납품했다.

그러자 D사는 계약서대로 인스턴스 40만 개를 라벨링 하면 이미지 4만 장을 모두 가공할 수 있는 것으로 인지했다며 나머지 1,900장에 대해서도 추가 작업을 요청했다. 심지어 무상으로 해달라며 대금 지급도 미루고 있다.

하지만 C사는 인스턴스 당 단가를 기준으로 계약했으므로 문제가 없다는 입장을 전달하고 무상 추가 작업을 거절했다. 만약 D사의 요구를 수용한다면 인스턴트 당 비용을 정산하는 라벨링 작업자에게 C사가 추가로 비용을 지급해야 해 경영상 손실을 입게 된다.

D사 주장

업계 기준이라고 해서 최종 결과물을 이미지가 아닌 인스턴스 수량으로 계약했습니다. 그런데 추정된 평균 인스턴스 가공량과 달랐다면 이미지를 적게 작업해 왔네요….

D사는 VR 여행 콘텐츠를 제작하기 위해 보유하고 있던 여행지 이미지 4만 장의 Bbox 라벨링을 C사에 의뢰했다. Bbox 라벨링은 이미지 개수가 아닌 인스턴스 개수를 기준으로 하는 것이 업계 관행이라고 해서 총 40만 개의 인스턴스를 작업하는 것으로 계약했다. 각 이미지 당 10개의 인스턴스를 라벨링 하면 될 것으로 판단했기 때문이다.

그런데 최종적으로 결과물을 받아 보니 4만 장에 못 미치는 3만 8,100장의 이미지만 작업이 되어 있었다. 이에 대해 D사에 문의하니 이미지 1장당 작업한 평균 인스턴스 수가 계약 전 추정된 10개가 아니라 10.5개였기 때문이라고 답변했다.

하지만 C사는 인스턴스 가공을 기준으로 계약하긴 했어도 원래 D사에 작업을 요청한 이미지는 총 4만 장이기 때문에 나머지 1,900장에 대해서도 추가 작업해 줄 것을 요청했다. 실제 이미지 당 평균 인스턴스가 10.5개라고는 하나 계약 전 추정된 10개와 차이가 크지 않고, 잔여 이미지 1,900장이 전체 수량의 5%도 되지 않기 때문에 무상으로 추가 작업을 한다고 해도 D사에는 큰 부담이 없을 것으로 판단했다. 라벨링 작업의 상당수가 오토 라벨링을 통해 쉽게 가능하다는 점도 염두에 둔 요구였다.

그러나 D사는 계약서가 인스턴스 수량을 기준으로 작성되었고 추가 작업을 진행하면 라벨링 작업자에게도 비용을 지급해야 한다며 무상 추가 작업을 거부했다. 작업자를 활용하지 않고 오토 라벨링이 가능한 부분이라도 추가 진행을 C사에 요청했지만, 이 또한 거부하며 조속한 대금 지급을 촉구하고 있다.

한편 대금 지급 지연과 관련해 D사는 의도적인 것이 아닌란 입장이다. 회계 절차상 발생한 단순 지연일 뿐이며 현재 대금 지급 절차를 진행 중이란 것이다.



알아두기

데이터 수량 관련 분쟁

이 사례와 같이 가공된 최종 데이터의 수량에 관한 분쟁을 피하기 위해서는 계약 이전에 기술 협상 시 최종 납품 수량, 대금 지급 일자, 추가 작업 조건 등을 구체적으로 합의하고 조율해 협의의 사항을 계약서나 부속서류 등에 기재할 필요가 있습니다.

01 데이터 생산

쟁점 3.

데이터 가공 시 계약 조건을 충족하기 힘든 경우 계약 위반 여부



E사와 F사는 키워드 속성별 라벨링 계약을 체결했다. E사는 라벨링 정확도 95%, 특정 알고리즘 사용을 계약 조건으로 걸었다. F사는 작업 중 특정 알고리즘만으로는 정확도를 충족할 수 없음을 발견하고 다른 알고리즘을 사용하여, 계약한 정확도(95%) 근사치의 결과물을 납품했다. E사는 이를 계약 위반을 주장했고, F사는 최적의 결과물이라고 응수해 분쟁이 발생했다.

E사 주장

데이터 라벨링을 의뢰하고 최종 결과물을 받았는데 계약 조건 2가지를 모두 만족하지 않습니다. 제공해 준 가이드라인으로는 최선의 결과물이라며 계약 위반을 인정하지 않네요….

E사는 소셜 미디어 알고리즘 전문 기업으로 SNS 게시물을 분석해 이용자들에게 적합한 콘텐츠를 추천하는 기술을 보유하고 있다. 이를 위해 연관어 및 트렌드 등을 분석하고 있는데, 최근 분석을 보다 정교하게 진행하기 위해 각 키워드별 감성 분석 모델을 구축하는 프로젝트를 시작했다.

이와 관련 데이터 라벨링 서비스 기업인 F사와 키워드 속성별 라벨링 계약을 체결했으며, 라벨링 결과물의 정확도가 95% 이상이어야 하고 반드시 특정 알고리즘을 활용해야 한다는 조건을 계약서에 명시했다.

E사는 F사가 소셜 데이터 라벨링 과정에서 여러 가지를 협의해 오는 등 열의를 보여 작업 결과물에 대한 기대가 컸다. 하지만 막상 최종 결과물을 받아

보니 라벨링 정확도는 92.4%에 불과했다. 계약한 95%를 충족하지 못한 것이다. 더욱이 특정 알고리즘만을 활용하기로 계약했음에도 일부 데이터는 다른 알고리즘을 통해 라벨링을 수행한 것이 확인되었다.

이에 E사는 F사가 계약 조건을 위반했다고 판단하고 계약 해지와 프로젝트 일정 지연에 대한 손해배상을 청구했다. 하지만 F사는 이를 받아들이지 않고 있다.

F사 주장

최적의 결과물 획득을 위해 일부 데이터를 계약과 다른 방식으로 작업했습니다. 수차례 가이드라인 변경 등으로 애로가 많았는데 무조건 계약 위반이라며 손해배상을 요구하네요….

F사는 E사와 키워드 속성별 라벨링 데이터 구축을 위한 계약을 체결하고 E사로부터 학습용 데이터와 라벨링 가이드라인을 전달받았다. 계약 조건은 라벨링 정확도 95%, 특정 알고리즘 사용이었다.

그런데 데이터 라벨링 작업을 진행하다 보니 전달받은 가이드라인이 실제 데이터에 적용하기는 모호한 부분이 많았다. 이에 E사와 협의해 여러 차례 가이드라인을 수정했고, 변경된 가이드라인에 따라 이미 라벨링을 완료한 데이터도 여러 차례 수정하는 등 작업을 반복해야 했다.

뿐만 아니라 가이드라인 변화에 따라 일부 데이터의 경우 계약 시 요구한 특정 알고리즘을 활용할 수 없음을 확인했다. 자체 테스트 결과 계속 특정 알고리즘을 활용할 경우 정확도가 떨어지는 문제가 발생했기에 전체 데이터의 절반가량은 더 적합한 다른 알고리즘을 활용해 작업했다.

물론 이는 계약 조건에 어긋나지만 또 다른 계약 조건인 정확도를 충족하려면 불가피한 조치였다. 무엇보다 라벨링 업무에서 가장 최우선 목표는 정확도 향상이라고 판단해 빠르게 다른 알고리즘으로 전환했기에 당초 계약한 95%에 근사한 92.4%의 정확도를 얻을 수 있었다.

하지만 E사는 서류상 명기된 계약 조건만을 제기하며 계약 해지를 요구하고 있다. 심지어 프로젝트 지연에 따른 손해배상까지 청구해 위약금 지급을 요구하고 있다.



알아두기

데이터 품질 관련 분쟁

이 사례와 같이 가공된 데이터의 품질에 관한 분쟁을 피하기 위해서는 특정 알고리즘 활용 여부, 가이드라인 변경 시 동반되는 이행 사항 변경 범위 등에 대한 구체적인 협의 후 계약 사항을 계약서에 명시하는 것이 좋습니다.

01 데이터 생산

쟁점 4.

데이터 생성 시 유사 데이터를 중복으로 판단하는 기준



G사와 H사는 반려동물 이미지 데이터 공급 계약을 맺었다. H사는 생성형 AI 기술로 이미지를 조합해 가공한 데이터를 전달했는데, 많은 수의 유사 이미지로 인해 G사의 이미지 인식 기술이 작동하지 않았다. G사는 이미지 중복 문제를 제기하며 추가 작업을 요구했고, H사는 생성형 AI 기술의 근원적 한계에 대해 계약 시 설명했다며 거부하면서 분쟁이 발생했다.

G사 주장

이미지 생성형 AI 회사를 통해 받은 데이터에 중복 이미지가 너무 많습니다. 그런데 이미지를 조합해 생성하는 데 따른 본질적인 한계라며 추가 작업을 거부하네요….

G사는 SNS에서 사람들의 반려동물을 애니메이션 캐릭터로 바꿔 주는 AI 앱을 개발하고 있었다. 최근 사람들이 키우는 동물이 개, 고양이 외에도 파충류, 곤충, 수상 생물 등으로 매우 다양하기에 수많은 종의 동물 이미지가 필요했다.

그러나 인터넷 등을 통해 수집할 수 있는 반려동물 이미지는 매우 한정적이고 해상도나 출처 또한 문제가 될 수 있어 G사는 데이터 확보에 어려움을 겪고 있었다. 이때 H사의 이미지 생성형 AI 서비스를 알게 되었다. 이 기술을 사용하면 각기 다른 특성을 가진 동물 이미지를 제한 없이 만들어 낼 수 있었다.

이에 G사는 H사와 반려동물 100종에 대해 각각 100건씩 총 1만 건의 데이터를 제공받는 계약을 체결하고 계약 기간은 3개월로 명시했다. 약속한 기한

이 도래해 전달받은 이미지들은 만족할 만한 수준으로 보였다.

하지만 육안으로 볼 때와 달리 AI가 생성한 이미지들에는 한계가 있었다. 동물의 눈, 코, 입 등 주요 부위는 동일하고 털의 색이나 길이 등만 변경한 유사 이미지들이 많아 G사가 개발한 이미지 인식 기술이 중복 이미지로 분석하는 바람에 테스트 앱이 제대로 작동하지 않은 것이다.

이에 대해 항의하자 H사는 AI로 반려동물 이미지를 생성할 때 조합을 활용하기 때문에 데이터의 양이 많아지면 유사한 데이터가 생성될 수밖에 없다고 설명했다. 가상 이미지를 생성할 때의 어쩔 수 없는 한계라는 것이다. 즉 적법한 방식으로 생성된 데이터라며 G사의 추가 데이터 생성 요청은 받아들이기 힘들다고 답변했다.

H사 주장

AI가 이미지를 생성하는 프로세스를 사전에 고지하고 계약을 이행했습니다. 그런데 일부 유사한 이미지를 데이터 중복이라며 추가 작업을 요구하네요….

H사는 이미지 생성형 AI 기술을 보유하고 있으며 상당한 양의 반려동물 이미지를 필요로 하는 G사와 100종, 1만 건의 데이터 계약을 맺었다. H사가 생성하는 데이터는 저작권자로부터 이용허락을 받아 저작권 문제가 해결된 데이터를 이용해 학습하여 생성된 것이었다.

계약에 따라 H사는 한 달 후 1차 데이터를 제공했고, G사는 이를 1주일간 검증한 후 나머지 데이터도 계약 기간 내 전달해 달라고 했다. 그런데 모든 데이터를 제공한 후 문제가 생겼다. G사가 일부 이미지들이 매우 유사하다며 2,000건을 추가 작업해 달라고 요구한 것이다.

하지만 AI가 조합해 이미지를 생성하는 데는 한계가 있기 때문에 정교한 이미지 인식 기술을 통해 분석할 경우 유사도가 높게 나올 수 있고 이는 사전에 설명이 된 부분이었다. 또한 G사가 추가 작업을 요구한 2,000건은 전체 데이터의 20%에 해당하는 큰 부분이었다. 이미 다른 클라이언트의 의뢰를 받아 가용할 수 있는 인력이 없는 H사로서는 감당하기 힘든 요구였고, 계약 이행상에도 문제가 없다고 판단해 G사의 요구를 받아들이지 않았다.



알아두기

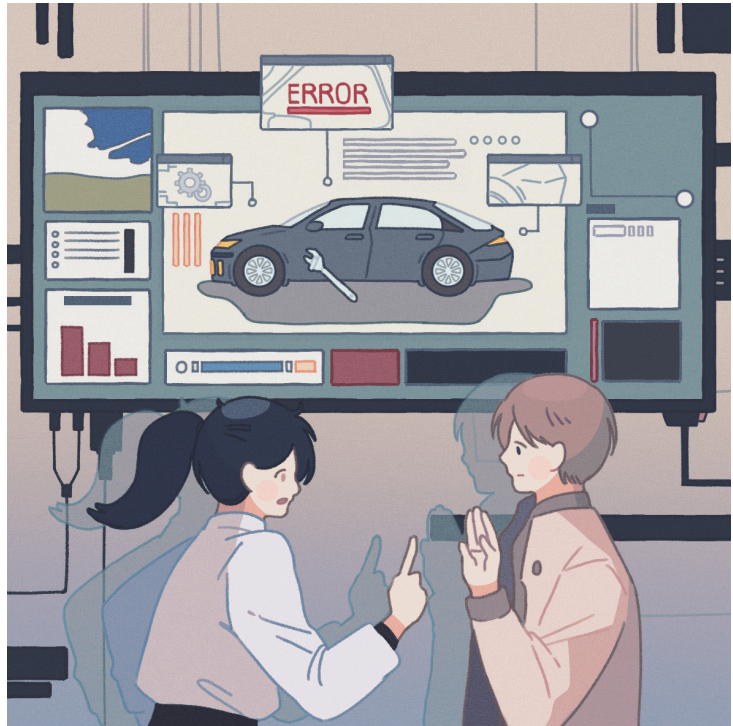
데이터 유사성 관련 분쟁

이 사례와 같이 가공된 데이터의 유사성(중복)에 관한 분쟁을 피하기 위해서는 계약 시 데이터의 품질 수준에 대해 계약서에 명시하고 하자 보수 작업에 대한 기준 및 작업 수량 등 구체적인 사항에 대해서도 협의가 필요합니다.

01 데이터 생산

쟁점 5.

데이터 납품 후 하자 보수의 범위



I사와 J사는 스마트 제조 시스템 데이터셋 및 학습 모델 납품 계약을 체결했다. I사는 막대한 비용을 들여 구축한 신규 시스템에서 데이터셋이 오류를 일으키며 작동하지 않자 하자 보수 이행을 요구했고, J사는 검사/검수 및 초기 작동 테스트에서 정상 작동이 확인되었으므로 납품한 데이터의 품질이나 학습 모델 문제가 아니라며 하자 보수를 거부해 분쟁이 발생했다.

I사 주장

데이터 오류로 새로 구축한 제조 시스템이 작동하지 않습니다. 이에 데이터셋과 학습 모델 구입 업체에 하자 보수 이행을 요구했지만 자신들의 문제가 아니라며 계속 회피하네요...

I사는 스마트 공장으로 전환하면서 기존 방식에서 벗어난 새로운 제조 시스템을 도입했다. 최신 기술이 접목된 머신러닝 기반의 자동차 모터용 부품 생산량 예측, 부품 제조 시의 수율 및 품질 향상을 위해 데이터 전문기업인 J사의 데이터셋과 데이터 학습 모델을 일괄 구매하는 계약을 체결한 것이다.

그런데 향후 J사로부터 납품받은 데이터에서 문제가 발견되었다. 기존 데이터와의 호환성, 시스템 확장성이 충족되지 않아 에러가 발생한 것이다. 이에 하자 보수 이행을 요구했지만 J사는 납품 당시 검사 및 검수 결과와 초기 작동 시 문제가 없었기에 이는 납품용 데이터의 품질 또는 학습용 모델 오류가 아니며, 시스템의 버전 문제이거나 현재 생산되는 데이터의 형식 변경 및 생산

주기의 비일관성 등이 오류의 원인이라고 밝혔다.

이러한 설명에 따라 I사는 J사가 지적한 항목을 변경하는 등 시스템에 대한 적절한 조치를 취한 후 다시 테스트를 실시했다. 그럼에도 이전과 동일한 오류가 다시 발견되자 I사는 J사에 상황을 설명하고 재차 하자 보수를 요구했지만 J사는 여전히 납득하지 않고 있다. 이를 하자 보수 이행에 관한 계약 불이행이라고 판단한 I사는 막대한 자본을 들여 구축한 새로운 제조 시스템이 데이터 문제로 사용 불가능한 데 따른 위약금을 J사에 청구했다.

J사 주장

정상적으로 데이터셋을 납품하고 검사/검수도 마쳤습니다. 그런데 데이터 형식 변경 등으로 발생한 오류까지 하자 보수를 요구해서 거부했더니 위약금을 청구하네요….

J사는 새로운 스마트 제조 시스템을 구축하는 I사에 데이터셋과 학습 모델을 판매했다. 당연히 납품 절차는 정상적으로 이루어졌다. 납품 당시 I사의 기준에 따라 검사/검수를 문제없이 통과했으며 I사의 시스템에서 초기에 정상적으로 작동함을 확인했다. 또한 I사의 핵심 요구 사항이던 자동차 모터용 부품 생산량의 실시간 업데이트, 예측 모니터링을 포함한 모터 제조 수율에 대한 결과도 모두 정상임을 확인했다.

이후 I사의 자체 검사/검수 과정에서 비슷한 오류가 몇 차례 발생하긴 했지만 이는 시스템의 버전 문제임을 상호 확인했다. 또한 I사가 최신 시스템에서 다양한 형식으로 신규 데이터를 생산할 때 발생할 수 있는 문제에 대한 가이드라인으로 데이터 활용 권고를 작성해 별도 제출했다.

그런데 I사는 지속된 데이터 오류 발생을 새로운 시스템 작동이 불가능하며 하자 보수를 요구하고 있다. 계약서의 '추가로 분석을 원하는 데이터에 대한 데이터 가공 서비스는 1회에 한하여 제공한다'라는 항목을 해당 문제와 결부시켜 위약금을 요구하고 있는 상황이다.

하지만 J사는 해당 계약 항목이 I사의 사업 확장과 관련해 모터용 부품 이외의 생산 품목에 대해 적용하는 것이란 입장이다. 이에 기납품한 데이터셋에 대해서는 하자 보수 의무가 없다고 판단해 I사의 위약금 청구에 대해 동의할 수 없다는 의견을 서면으로 전달했다.



알아두기

데이터 하자 보수 관련 분쟁

이 사례와 같이 데이터의 하자 보수에 관한 분쟁을 피하기 위해서는 발생 가능한 하자의 내용, 범위 등에 대한 명확한 기준을 협의하고 계약서에 하자 보수의 기준과 방법, 시기 등을 명시할 필요가 있습니다.

01 데이터 생산

쟁점 6.

데이터 검증 시 주관적 판단 개입의 기준



K사와 L사는 어린이 교육용 대화 데이터셋 및 대화 느낌 분류 계약을 체결했다. K사는 결과를 검증 후 일반적인 대화형 AI 챗봇 수준의 전개보다 완성도가 떨어지는 결과물에 대해 전체 재작업을 요구했고, L사는 1차 답변으로 충분한 정보 전달이 되었다면 추가 질문이 필요하지 않으며, 이는 가이드라인 이상의 주관적 품질 검증이라며 재작업을 거부해 분쟁이 발생했다.

K사 주장

대화 데이터셋 구축을 의뢰했는데 검증해 보니 일반 챗봇의 대화 구성과 달리 1차 정보 전달만으로 대화가 거의 종료됩니다. 이에 저품질에 따른 재작업을 요청했는데 안 된다네요….

K사는 어린이 교육용 대화형 AI 서비스를 개발하면서 대화(인터랙션) 기능 고도화를 위해 L사와 대화 데이터셋 구축과 대화에 대한 느낌 분류 작업 계약을 체결했다. 어린이들의 경우 대화의 내용 못지않게 말투나 단어 선택 등이 중요하기에 K사는 L사에 특정 대화 주제를 정리해 사전에 전달했다.

다만 사용하는 단어나 문장 형태 등 대화 생성 및 느낌(감정)에 대한 분류는 작업자에 따라 달라질 수 있고, 그 결과로 데이터셋의 품질 검증에 대해서도 주관적인 개입이 발생하기 때문에 K사는 전체 작업 전 L사에게 초기 샘플 데이터를 요청했다. 이 초기 데이터셋은 대화가 전반적으로 자연스럽게 어린이들도 쉽게 이해할 수 있는 대화 구성이었다.

그런데 L사로부터 최종 납품받은 데이터셋을 사내에서 자체 검증하는 과정에서 문제가 발견되었다. 대화 데이터셋의 주관적 요소가 많아 품질 검증에 시일이 많이 소요되었을 뿐만 아니라 일부 데이터셋은 매우 부자연스러워 사용이 힘든 것으로 판단된 것이다.

특히 일반적인 대화형 챗봇 서비스와 달리 질문에 일방적인 정보를 전달하는 1차 답변만으로 대화를 종료하는 등 학습에서 가장 중요한 인터랙션이 거의 없었다. 이에 K사는 데이터 전체에 대해 재작업을 요청했지만, L사는 해당 기준이 지나치게 주관적이라는 이유로 재작업 요청을 거부했다.

L사 주장

주관적 요소가 큰 대화 데이터의 특성을 보완하기 위해 샘플 데이터셋 검증 완료 후 작업을 진행했습니다. 그런데 초기 가이드 이상을 요구하며 전체 재작업을 요구하네요….

L사는 K사로부터 대화 데이터셋 구축을 의뢰받는 과정에서 수집 및 검증에 주관적 판단이 들어갈 수 있는 대화 데이터의 특성을 보완하기 위해 초기 수집한 데이터의 샘플 데이터셋을 K사에 보내 검수받았다.

그리고 검증 완료된 샘플 데이터셋의 데이터를 수집한 작업 인력을 대상으로 추가 교육을 진행해 전체 데이터를 수집했다. 또한 대화 데이터셋의 품질을 제고하기 위해 작업자의 경우 유사 프로젝트 경험이 많은 사람을 처음부터 선발해 교육하고 프로젝트 종료까지 유지했다.

하지만 K사는 질문에 대한 정보 전달 차원의 1차 답변으로는 대화의 완성도가 떨어진다고 하며 답변 이해 여부 확인, 추가 피드백 등을 대화에 추가할 것을 요청하며 전체 데이터의 재작업을 요구했다.

이에 L사는 1차 답변에서 정보를 깊게 안내하는 것으로도 자연스러운 대화라고 볼 수 있음을 설명하고, 동시에 K사가 제공한 가이드 및 기준 범위 내에서는 최적의 데이터를 수집했다고 답변했다. 또한 L사가 대화 진행 단계에서 원하는 바를 가이드라인에 구체화했다면 작업 시 반영했을 것이므로, 기준을 넘어서 지나치게 주관적인 품질 검증 결과로 인한 재작업 요청은 수긍하기 어렵다는 의견을 전달했다.



알아두기

데이터 검증 관련 분쟁

이 사례와 같이 주관적 개입이 있는 데이터의 검증에 관한 분쟁을 피하기 위해서는 데이터의 객관적 품질 기준을 명확하게 제시하고 기준에 부합하지 않는 경우 추가 작업 관련 범위도 계약서에 명시하는 것이 좋습니다.

02 데이터 거래

쟁점 7.

데이터 납품 시 사용된 데이터 가공·생성 모델의 소유 기준



M사와 N사는 M사의 온라인 쇼핑몰 플랫폼 고도화를 위한 데이터 가공 계약을 체결했다. 납품 후 M사는 내부 개발자의 지속적인 기능 업데이트를 위해 커스터마이징 시 데이터 가공·생성 모델 탑재를 요구했고, N사는 데이터 가공·생성 모델의 경우 계약 내용에 포함되지 않는 고유의 지식재산이라며 제공 불가능하다고 밝혀 분쟁이 발생했다.

M사 주장

플랫폼 고도화를 위해 데이터 가공 계약을 체결했는데 당사 커스터마이징 데이터 가공·생성 모델이 누락되었습니다. 하지만 데이터만 납품 대상일 뿐 커스터마이징 모델은 제공할 수 없네요...

M사는 온라인 쇼핑몰 플랫폼을 운영하면서 ICT 기반 고객 관계 관리(CRM) 시스템을 통해 회원사들을 관리하고 있었다. 지난 5년간 플랫폼을 운영하며 쌓인 데이터를 보다 효율적으로 분석, 관리하기 위해 M사는 N사와 AI 기술을 활용한 데이터 가공 계약을 체결하고 플랫폼 고도화 작업을 진행했다. 그 결과 M사가 회원사들을 더 잘 관리할 수 있게 되었을 뿐 아니라 회원사들도 자신들의 고객 관리를 보다 편리하게 할 수 있게 되었다.

그런데 이후 회원사들의 요청에 따라 플랫폼의 일부 기능을 업그레이드하려고 데이터를 가공하려 하자 M사의 내부 개발자가 활용할 수 있는 커스터마이징 데이터 가공·생성 모델이 포함되지 않았음을 알게 되었다. 계약 과정에서 M

사는 자사의 플랫폼 특성과 운영 메커니즘을 N사에 상세히 공유하면서, 데이터 뿐 아니라 추가 수집되는 데이터들을 기존 플랫폼상에서 처리할 수 있도록 데이터 가공·생성 모델이 필수 포함되어야 함을 강조했는데 누락된 것이다.

추가 데이터 가공이 되지 않으면 지속적인 서비스 유지가 어렵다고 생각한 M사는 N사에 데이터 가공·생성 모델을 제공해 줄 것을 요구했다. 해당 모델이 양사 간 계약에 의해 M사의 플랫폼에 커스터마이징된 것이기에 다른 곳에서는 사용이 불가능하므로 M사가 플랫폼에 탑재해도 문제가 없다고 판단했다. 하지만 N사는 데이터 가공·생성 모델의 제공은 추가적인 계약 사항이 필요한 사항이라며 거절했다.

N사 주장

온라인 쇼핑몰 플랫폼 회사와 데이터 가공 계약을 맺고 커스터마이징 모델을 활용해 납품을 완료했습니다. 그런데 자신들의 플랫폼에 탑재할 데이터 가공·생성 모델까지 달라고 하네요….

N사는 M사의 플랫폼 고도화를 위한 데이터 가공 계약을 맺고 자체 솔루션을 활용해 M사 커스터마이징 모델을 만들어 데이터를 가공, 납품을 완료했다. 그런데 M사가 회원사 요구 수용을 위한 플랫폼 일부 수정 중 자신들의 개발자가 고객 관리를 위해 데이터를 가공할 수 있는 커스터마이징 모델이 누락되었다며 연락해 왔다.

하지만 N사는 M사와의 계약은 데이터 가공·생성 모델이 데이터 그 자체만 포함되는 것으로 해석했다. 데이터 납품 이후 3년간 무상 업데이트 지원에 대한 내용이 계약에 포함되어 있지만 이것이 데이터 가공·생성 모델 제공은 아니기 때문이다.

그래서 M사에 3년간 자신들이 제공하는 무상 지원 서비스만으로 플랫폼 운영에는 문제가 없을 것이라고 설명했다. 또한 M사가 요구하는 데이터 가공·생성 모델은 N사의 핵심 기술인 AI 솔루션을 기반으로 커스터마이징한 것이므로 궁극적으로는 지식재산권의 범위에 있다고 부연했다.

그리고 만약 M사가 데이터 가공·생성 모델을 플랫폼에 탑재하면서 자체적으로 지속적인 기능 개선 및 원활한 서비스 운영을 하고 싶다면 추가적인 계약을 통해 모델의 사용권을 획득해야 한다는 의견을 전달했다. 즉 N사는 M사와의 계약은 데이터의 납품에 대한 계약으로 데이터 가공·생성 모델까지 제공해야 한다는 M사의 주장에 동의할 수 없다는 입장이다.



알아두기

데이터 납품 관련 분쟁

이 사례와 같이 데이터 납품에 관한 분쟁을 피하기 위해서는 계약 과정에서 거래의 대상이 무엇인지, 데이터인지 그 외 솔루션이나 모델 등을 포함하는지를 구체적으로 협의하고, 협의 내용을 계약서에 명시할 필요가 있습니다.

02 데이터 거래

쟁점 8.

데이터 거래 시 계약 후 납기 변경과 관련한 기준



O사와 P사는 AI 맞춤형 안경 솔루션 개발을 위한 데이터 수집 및 가공 계약을 체결했다. O사는 자사 담당자의 퇴사 후 프로젝트 지연 사실을 알게 되어 P사에 계약서의 지체상금 항목에 대한 이행을 요구했고, 이에 대해 P사가 퇴사한 담당자와 납품 지연 및 지체상금 미지급에 대한 구두 약속을 받았다고 응답하면서 분쟁이 발생했다.

O사 주장

데이터 수집 및 라벨링 프로젝트 마감 일주일 전 납품 지연 예정 사실을 알게 됐습니다. 데이터 업체는 퇴사한 자사 담당자와 지연 협의가 되었다며 지체상금 지불을 못한다네요….

O사는 AI 맞춤형 안경 솔루션을 개발하는 업체로, 한국인의 얼굴형과 눈, 코, 입 위치 등을 고려해 온라인으로 안경테를 추천하는 개인 맞춤형 서비스를 개발 중이다. AI 모델 학습 고도화를 위해 다량의 이미지가 필요했기에 P사를 이미지 수집 및 가공업체로 선정해 90일 후 데이터를 납품받기로 계약을 체결했다. 또 완성된 솔루션은 오랜 기간 거래를 이어온 고객사에 판매하기로 하고 2개월 후 납품하는 공급 계약을 별도 체결했다.

그런데 약속한 납품 완료일을 일주일 앞두고 P사와 업무를 진행하던 O사의 담당자가 개인적인 사유로 갑자기 퇴사하게 되었다. 그리고 인수인계 후 O사의 새로운 담당자는 P사의 납품 지연으로 인해 전체 프로젝트의 일정이 지연

될 상황임을 알게 되었다.

O사는 이미 자신의 고객사와 솔루션 공급 기한이 포함된 계약을 체결한 상태여서 당황할 수밖에 없었다. 고객사와의 계약에도 P사와의 계약과 마찬가지로 지체상금 조항이 있어 금전적 손실을 보아야 하는 상황이었다. 사실 그보다 더 큰 문제는 오랜 고객사와의 신뢰가 깨지는 것이었다. 시장 내에서 그동안 다져온 O사의 이미지가 실추되는 것도 피할 수 없었다.

이에 O사는 P사에게 지연일수에 해당하는 지체상금을 지불하라고 통보했다. 그러나 P사는 퇴사한 기존 담당자와 구두상으로 이미 산출물 납품 지연 통보에 대해 협의했으므로 지체상금에 대해서도 양해가 된 것으로 안다며 지체상금을 받아들이지 않고 있다.

P사 주장

데이터 수집의 어려운 상황을 설명하고 지체상금 미지급에 대한 구두 양해를 받았 습니다. 하지만 해당 담당자 퇴사 후 새로운 담당자는 무조건 계약서대로 지체상 금을 입금하라고 하네요….

P사는 데이터 라벨링 전문 기업으로 AI 맞춤형 안경 솔루션을 개발 중인 O사와 사람의 얼굴형, 눈과 코, 귀의 위치 등에 대한 이미지를 수집하고 세그멘테이션 라벨링을 하는 작업을 계약했다. 당초 총계약 금액 5,000만 원, 계약 기간 3개월로 지체상금이 없는 표준계약서 양식을 제안했으나, O사의 법무팀이 지체상금 항목 추가 의견을 제시해 최종적으로는 O사가 새로 작성한 계약서로 계약을 체결했다.

데이터 수집은 온라인으로 크라우드 워커를 모집해 진행했는데 예상보다 신청자가 적었다. 이에 P사는 O사와 계약한 납품 완료 기간을 맞추지 못할 것 이라 판단하고 일정 지연 협의를 요청했다. O사의 담당자는 본인들이 자체적 으로 데이터를 수집할 때도 비슷한 어려움이 있어 충분히 이해한다며 최대한 빨리 작업을 완료해 달라고만 당부했다. 또한 지체상금에 대해서도 자사에 잘 이야기해 지불하지 않는 걸로 해주겠다고 구두로 약속했다.

하지만 프로젝트 종료 일주일 전 O사의 해당 담당자가 갑자기 퇴사를 해버 렸다. 이후 인수인계를 받은 새로운 담당자는 납품 지연을 문제 삼으며 계약서 대로 지체상금을 지불할 것을 요구해 왔다.

이에 P사는 O사의 기존 담당자와 한 약속을 상기시키며 지체상금을 지불할 수 없다고 답변했다. 하지만 O사는 퇴사한 담당자와 한 약속일 뿐이라며 양사 의 계약에서 더 우선되는 것은 문서로 작성된 계약서라고 강조했다. 그리고 P 사가 통보받은 날짜까지 지체상금을 입금하지 않자, 퇴사한 담당자의 구두 약속에 대한 도의적 책임도 지지 않고 즉시 입금하지 않으면 내용증명을 보내겠다고 엄포를 놓고 있다.



알아두기

데이터 납기 관련 분쟁

이 사례와 같이 데이터의 납기 일에 관한 분쟁을 피하기 위해서는 계약에 납기 일자를 명시 하는 것뿐 아니라, 납기 변경이 필요한 경우 계약 종료 전 언제 까지 협의를 해야 한다는 내용 과 계약 불이행 시 손해배상 등 을 구체적으로 협의하고, 협의 내용을 계약서에 명시할 필요가 있습니다.

03 데이터 활용

쟁점 9.

데이터 납품 후 데이터의 폐기 및 활용 기준



Q사와 R사는 Q사의 콜센터 기능 업그레이드를 위한 고객 상담 데이터 라벨링 계약을 체결했다. Q사는 납품 후 R사가 자사 데이터를 즉시 폐기하지 않고 내부 데이터로 활용한 사실을 발견하고 계약 위반을 주장했고, R사는 폐기 지연은 피드백 대비를 위한 것이었고 Q사가 자신들의 데이터라고 오인한 것도 오픈 플랫폼에서 구한 데이터라고 주장해 분쟁이 발생했다.

Q사 주장

데이터 라벨링을 의뢰하면서 자사 데이터를 제공했습니다. 프로젝트 종료 후 해당 데이터를 데이터 업체가 홍보 및 영업에 임의 사용한 사실을 발견했는데 인정하지 않네요….

Q사는 B2C 거래를 메인 비즈니스로 하는 통신판매 회사로 콜센터의 기능 업그레이드를 위해 ARS 및 AI 챗봇 고도화를 추진하고 있었다. 자사 개발팀이 있었지만 보다 기능적으로 우수한 솔루션을 갖추기 위해 R사와 텍스트 데이터 라벨링 계약을 체결하고, 그동안 Q사가 수집한 고객 상담 데이터를 R사에 제공했다.

이후 납품은 계약대로 차질 없이 진행되었으나, 몇 개월 후 개발팀 담당자가 콜센터 관련 콘퍼런스를 다녀오면서 문제가 발생했다. R사가 자사 데이터를 AI 학습을 위해 임의로 사용해 포트폴리오를 작성한 것은 물론 경쟁사 영업 예까지 활용한 것으로 추정되는 내용을 확인한 것이다.

Q사는 R사가 데이터 납품 완료 후 계약 조건에 따라 납품한 데이터를 모두 폐기했을 거라 생각했다. 하지만 여러 루트를 통해 확인해 보니 R사는 Q사의 계약 목적, 즉 Q사의 데이터 가공 및 납품을 위해서만 활용되어야 하는 데이터로 AI 모델링을 진행해 홍보하고 있었다. 더욱이 해당 내용을 토대로 타 업체와 계약을 체결해 수익도 내고 있었다. 이와 관련 Q사가 R사 담당자와 통화했을 때도 해당 사실을 인정하는 취지의 발언을 들을 수 있었다.

이에 Q사는 R사의 계약 불이행에 따른 보상은 물론 자신들이 제공한 데이터를 학습한 AI 모델 홍보를 통해 R사가 타사와 체결한 계약 매출의 수익 배분 등을 요구했다. 또한 일련의 과정에 대해 법적 검토를 한 후 손해배상 등 민형사 조치도 진행하고 있다.

R사 주장

데이터 납품 후 해당 데이터가 계약 외 다른 목적으로 사용되었다며 클라이언트가 문제를 제기했습니다. 피드백을 위해 폐기를 지명한 것일 뿐인데 손해배상까지 요구하네요….

R사는 데이터 가공 기업으로 Q사가 제공한 고객 상담 데이터를 가공하는 계약을 맺고 납품을 완료했다. 그런데 몇 달 후 Q사가 계약 위반을 주장하며 문제를 제기해 왔다. 자신들이 제공한 데이터가 R사의 홍보 및 타사 영업에 사용되고 있다는 것이다.

하지만 R사는 “콜센터 콘퍼런스에서 발표한 프레젠테이션과 영상에 포함된 AI 모델링 관련 데이터는 Q사가 제공한 데이터를 학습한 것이 아니라 적법한 방법으로 오픈 플랫폼에서 구한 것”이라고 말했다.

또한 Q사가 R사 담당자에게 유선으로 해당 내용을 확인해 왔을 때 Q사의 데이터라고 인정했다는 오해에 대해서는, 데이터 활용이 아니라 납품 후 데이터 미폐기에 대한 인정이었다고 설명했다. 그리고 납품 후 데이터를 즉시 폐기하지 않은 이유에 대해서도 일반적으로 수정 요청이나 피드백을 고려해 얼마간 데이터를 보관했다가 폐기하고 있다고 설명했다.

실제로 R사는 Q사의 문제 제기 이후 즉시 데이터를 전량 폐기했다. 따라서 이는 계약 위반이 아니라 폐기 진행이 약간 지연된 수준으로 이에 대해 계약 위반이나 손해배상을 적용하기는 힘들다는 입장이다.



알아두기

데이터 폐기 관련 분쟁

이 사례와 같이 데이터 폐기와 관련한 분쟁을 피하기 위해서는 폐기 기간, 계약 이행 후 수집 및 가공한 데이터에 대한 폐기, 해당 데이터를 폐기했음을 확인할 수 있는 증빙 자료 제출, 불이행 시 손해배상 등을 계약서상에 명시할 필요가 있습니다.

03 데이터 활용

쟁점 10.

데이터의 재가공 및 활용 기준



S사와 T사는 G사의 챗GPT 교육 콘텐츠를 제작을 위한 데이터셋 제공 계약을 체결했다. S사는 T사로부터 임차한 데이터셋을 활용해 G사에 판매한 후 사용계약 기간 내 해당 데이터를 재가공해 L사에 추가 판매했고, 이를 발견한 T사가 S사에 데이터 임의 재가공 및 활용에 대해 계약 위반을 주장하며 분쟁이 발생했다.

S사 주장

T사와 데이터셋 제공 계약을 체결하여 G사에 콘텐츠를 제작, 판매한 후 계약 기간 내에 데이터셋을 재가공해 L사에도 판매했습니다. 완전히 새로 가공한 데이터인데 임의 사용 문제를 제기하네요….

S사는 직원용 챗GPT 활용 교육 콘텐츠를 제작해 G사에 판매하기 위해 실제 질의응답 시나리오를 기반으로 한 데이터셋이 필요했다. 여러 업체를 검토해 본 결과 T사의 데이터셋이 가장 적합하다고 판단해 구매하기로 내부 결재도 마쳤으나, S사의 담당자는 T사가 제안한 가격이 G사와 계약한 프로젝트 예산 범위를 초과하는 수준이라고 판단했다. 이에 비용 부담을 덜 수 있는 방법을 T사에 문의했고, T사는 데이터를 구매하지 않고 제공 계약으로 진행할 수 있다며 대안을 제시했다.

이를 받아들여 S사는 T사와 데이터 제공 계약을 체결했다. T사가 제시한 계약 내용에 데이터 활용과 관련한 몇 가지 조건이 걸려 있었으나 대수롭지 않

은 조항이라고 판단했다. 그리고 T사로부터 데이터를 납품받아 S사에 콘텐츠 판매를 성료했다.

얼마 후 S사는 L사로부터 또다시 챗GPT 교육 콘텐츠 제작 및 판매 의뢰를 받았고, 기존에 T사에서 임대한 데이터의 사용 기한이 아직 남았기에 그것을 재가공해 새로운 콘텐츠 판매도 성공적으로 마칠 수 있었다.

그런데 이 사실을 알게 된 T사가 S사에 데이터 무단 사용에 대한 이의를 제기해 왔다. 계약 위반에 따른 피해 보상 및 법적 대응까지 거론하고 있어 담당자도 많이 당황한 상태다.

물론 임대한 데이터를 최초 콘텐츠 판매 대상이었던 S사 관련해서만 활용한다는 조항이 있었으나, S사는 해당 데이터를 그대로 사용한 것이 아니고 별도의 공수를 들여 아예 새로운 데이터로 가공해 콘텐츠를 제작한 것이므로 문제가 없다고 판단했다.

게다가 S사가 S사가 아닌 L사에 새로 콘텐츠를 판매한 사실은 내부 정보인데 T사가 그것을 곧바로 알고 이의를 제기한 부분도 이상하게 생각하고 있다. S사에 제공한 데이터에 T사가 불법 소프트웨어를 설치한 건 아닌지 의심되는 상황이다.

T사 주장

데이터셋 임대 시 활용 기간과 대상을 명시한 계약서를 작성했는데 임의 재가공 후 판매한 것을 확인했습니다. 계약 위반 사실 확인을 요구했더니 아예 연락을 피하네요...

T사는 챗GPT 교육 콘텐츠를 제작해 판매하는 S사와 데이터셋 제공 계약을 맺었다. S사가 가격에 부담을 느끼고 있어, 데이터셋을 판매하는 대신 콘텐츠 제작 기간인 1년간만 제공하고 폐기하는 조건으로 계약서를 작성했다. 단 해당 데이터셋은 S사가 수주받은 S사의 콘텐츠 제작에만 활용한다는 단서를 달았다.

이후 S사는 T사의 데이터셋을 활용해 7개월 만에 콘텐츠를 성공적으로 제작해 판매했다며 감사 인사까지 전해 왔다. 그런데 이후 T사가 S사와 마찬가지로 직원용 챗GPT 교육 콘텐츠를 제작 중인 L사와 미팅하는 과정에서 황당한 사실을 알게 되었다. T사가 소개한 데이터셋에 대해 L사가 얼마 전에 이미 구매한 것과 동일하다고 얘기한 것이다.

세부 데이터를 확인해 보니 해당 데이터셋은 T사가 S사에 제공한 데이터셋을 활용해 재가공한 것이었다. 즉 S사는 T사와 S사와의 거래 시에만 데이터셋을 활용하기로 계약해 놓고, 그것을 임의 가공해 L사에 추가 판매한 것이다.

이에 T사는 S사에 사실 확인을 요구했으나, S사 담당자는 다시 연락을 주겠다고만 하고 이후 접촉을 피하고 있는 상황이다. 이에 T사는 피해 보상 규모를 산정하고 법적 대응을 준비하고 있다.



알아두기

데이터 제공 관련 분쟁

이 사례와 같이 데이터 제공에 관한 분쟁을 피하기 위해서는 계약서상에 명시된 제한된 목적으로만 데이터를 활용할 수 있고 그 외에 재가공 등 다른 목적으로의 데이터 활용은 금지한다는 조항을 계약서에 명기할 필요가 있습니다.

부록

01 데이터 분쟁조정 Q&A

01 데이터 분쟁조정 Q&A

Q1. 데이터분쟁조정위원회는 어떤 일을 하는 곳인가요?



A1. 데이터의 생산, 활용, 거래와 관련하여 분쟁이 발생하였을 때 소송까지 가지 않고 분쟁 해결을 도와드립니다. 이는, 재판상 화해와 동일한 효력을 갖게 됩니다.

Q2. 소송보다 분쟁조정이 좋은 점은 무엇인가요?



A2. 상대방과 원만한 해결을 할 수 있도록 지원해드리는 제도입니다.

소송 비용과 시간을 절약할 수 있으며, 데이터 전문가와 함께 해결 방안을 찾아볼 수 있습니다.

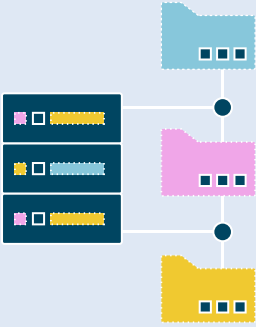
Q3. 데이터분쟁조정은 얼마나 걸리나요? 비용은요?



A3. 신청서가 접수된 날부터 45일 이내에 해결책(조정안)을 받을 수 있으며, 당사자 모두 동의할 경우 분쟁은 해결(조정성립) 됩니다.

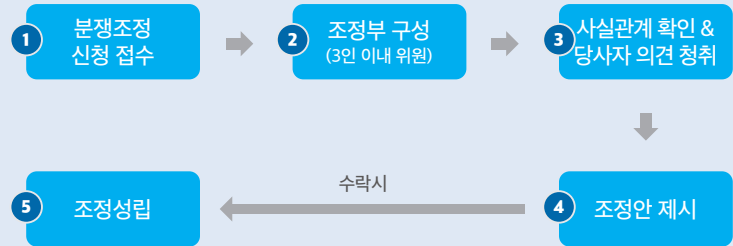
비용은 무료입니다.

(단, 위원회에서 조정비용을 부담하도록 정한 사건은 제외)



Q4. 데이터 분쟁 조정 절차는 어떻게 되나요?

A4.



Q5. 데이터 분쟁조정은 누가 해주나요?

A5.

위원회는 법조계, 학계, 공공, 산업계 전문가로 구성되어 있으며, 신청 내용에 따라 3명 이내의 위원으로 구성된 조정부에서 데이터 분쟁조정을 수행합니다.



Q6. 데이터 분쟁조정이 성립된 이후, 법원에 소송할 수 있나요?

A6.

조정조서는 재판상 화해와 동일한 효력을 갖게 되어 일반적으로는 동일한 소송은 제한됩니다



Q7.
데이터분쟁 조정안은 판결문처럼 무조건 따라야 하나요?

A7.
위원회에서 조정하는 **조정안(해결방안)**에
당사자 **모두가 동의**해야
판결문과 같은 효력이 부여됩니다.



Q8.
상대방이 분쟁조정에 응하지 않아도 분쟁조정은 성립할 수 있나요?

A9.
분쟁조정은 당사자의 상호 타협과 양보를 통한
해결 방안을 제시하고 수락하는 것을 **전제**로 하기 때문에,
상대방이 **거부**할 경우 **분쟁조정을 진행할 수 없습니다.**



Q9.
분쟁조정 신청을 하면, 신청인이 해야 할 일은 무엇인가요?

A9.
조정부는 **사실확인**을 위해 당사자 **의견 청취**,
현장 답사를 진행할 수 있으며,
필요시 조정 회의에 참석을 요청할 수 있습니다.



Q10.
구체적인 데이터 분쟁 사례를 알 수 있을까요?

A10.
데이터 생산·거래·활용에 관한 분쟁으로
다음의 경우에 해당됩니다.

데이터의 생산, 가공 등이 계약과 다른 경우

납기 지연, 비용 미지급 등 계약이 이행되지 않은 경우

계약과 다른 목적으로 데이터를 사용한 경우 등

⋮



Q11.
데이터 분쟁조정은 어디서 신청하나요?

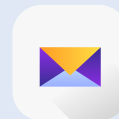
A11.
홈페이지(ddrc.go.kr)에서 회원가입 후
신청해주시면 됩니다.

Q12.
데이터 분쟁조정 관련 문의는 어디로 하나요?

A12.
데이터분쟁조정위원회 지원반(NIA)으로 연락주시면 됩니다.



053-230-4202
4210
4268



ddrc@nia.or.kr

알기쉬운 데이터 분쟁조정 제도 안내

인쇄 2024년 2월

발행 2024년 2월

발행처 데이터분쟁조정위원회(ddrc.go.kr)

© 데이터분쟁조정위원회, 2024

이 안내서는 저작권법에 따라 보호를 받는 저작물이므로
무단 전재 및 재가공을 금지합니다.

이 안내서에 실린 모든 내용(디자인, 이미지, 편집 구성 등)의
저작권은 데이터분쟁조정위원회에 있습니다.

이 안내서(PDF)는 자유롭게 배포 가능하나, 내용의 전부 또는 일부를
다른 용도로 이용하려면 반드시 저작권자의 서면 동의를 받아야 합니다.



과학기술정보통신부 데이터분쟁조정위원회 NIA 한국지능정보사회진흥원