

AI 시대의 변화, 기회로 전환하다

작년 말 시작된 오픈AI의 대화형 AI 서비스 '챗GPT'를 비롯해
기업들이 앞다퉈 선보이는 AI 기술은 따라잡기 벅할 정도로 발전을 거듭하고 있다.

AI 기술 경쟁에서 우위를 차지하려는 기업들의 움직임이 거세고
산업별로 AI 기술의 도입 수요도 급증하고 있지만, 전망에 대한 우려 섞인 시각도 만만치 않다.

급격히 발전하는 AI 기술이 사회 전반에 미치는 영향력과 산업계의 움직임,
그리고 비즈니스에 AI 도입을 원하는 기업들을 위한 전략을 소개한다.



AI가 바꿔놓을 삶과 산업의 미래

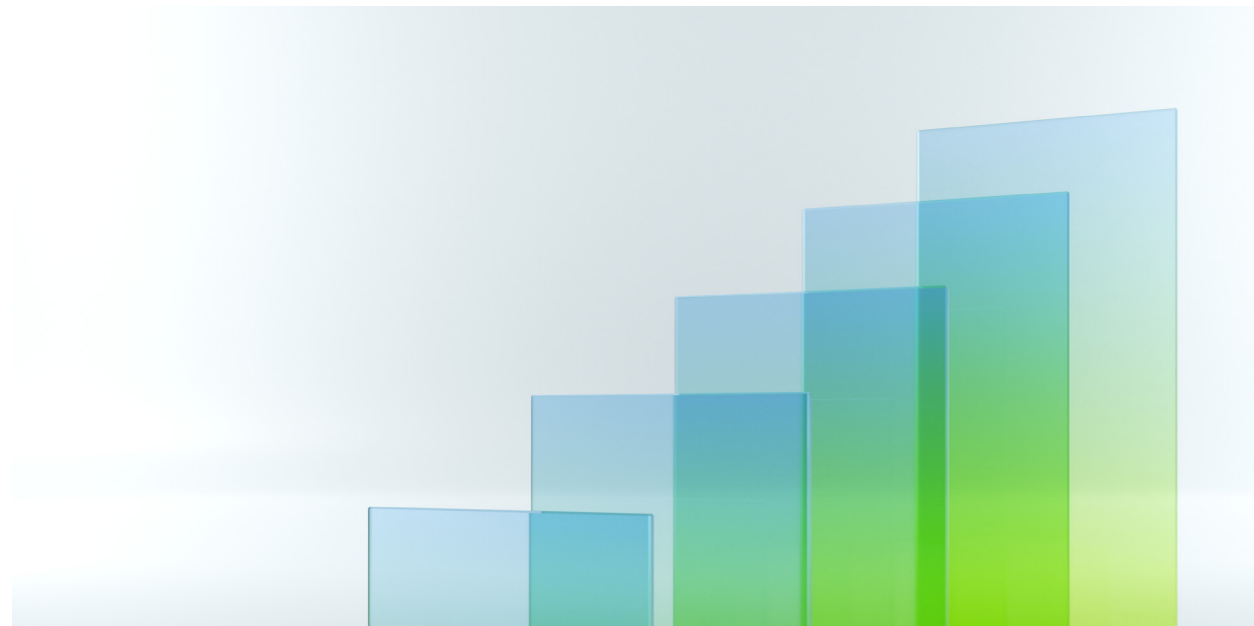
김우용 / 지디넷코리아 기자

오픈AI의 대화형 AI 서비스 ‘챗GPT’는 세상을 순식간에 태풍의 눈 속으로 몰아넣었다. 창의적이고 전문적인 작업, AI로 대체하기 힘들 것으로 여겼던 작업을 AI로 쉽게 대체하는 세상이 눈앞에 성큼 다가온 것이다.

올해 1월부터 3월까지 마이크로소프트, 구글, 오픈AI, 아마존웹서비스(AWS), 메타 등이 경쟁적으로 선보인 생성 AI 기술은 눈부신 발전 속도를 보였다.

미드저니나 스테빌리티AI의 스테이블디퓨전 같은 이미지 생성 AI는 어떨까. 사진인지 그림인지 구분하기 힘들 정도로 정교한 이미지를 순식간에 만들어낸다. 런어웨이는 이미지나 명령어 입력으로 새로운 동영상을 만드는 모델을 제공하고, 주크박스는 음악을 만들어준다. 빅테크 기업 뿐 아니라 전문 스타트업까지 성과가 눈부시다.

이처럼 곳곳에서 단기간에 쏟아내는 정보가 너무 많다 보니, 따라잡기 벅찰 수준이다. 세상이 천지개벽하는 착시마저 일으킨다.



급격히 발전하는 AI 기술

챗GPT는 작년 11월말 처음 등장했을 때만 해도 일부 전문가와 얼리어답터만의 관심거리였다. 그러나 오픈AI의 제어기술을 무력화하는 ‘챗GPT 해킹’이 유행처럼 번지고, 각종 시험에서 챗GPT가 통과하는 등 일반인의 관심영역으로 이동하는 데 그리 오래 걸리지 않았다. 오픈AI는 단순히 범용 대형언어모델(LLM)인 GPT-3.5의 특정 목적에서 활용성을 테스트하기 위해 챗GPT를 선보였다가 뜻밖의 인기를 얻게 됐다. 챗GPT는 두 달 만에 월간실사용자(MAU) 1억 명을 돌파하며 역사상 가장 빨리 성장한 서비스 자리에 등극했다.

1월 들어 오픈AI에 자본금과 인프라를 투자해온 마이크로소프트의 행보가 본격화됐다. 마이크로소프트는 오픈AI의 GPT 모델과 코덱스, DALL-E, 챗GPT 등의 특정 모델을 API로 활용할 수 있는 ‘애저 오픈AI 서비스’를 먼저 선보였다. 그리고 인터넷 서비스 ‘빙’에 GPT-4를 결합한 대화형 AI 기능을 탑재해 ‘뉴 빙’이란 이름으로 공개했다.

그리고 지난 3월, 마이크로소프트는 워드, 엑셀, 파워포인트 등에 적용한 생성 AI 기술을 시연했다. 전 세계에 온라인 이벤트로 생중계된 이 행사에서 공개된 ‘마이크로소프트365 코파일럿’ 시연 영상은 스티브 잡스의 아이폰 발표에 버금갈 충격을 줬다.

대화하듯 문서와 파워포인트 프리젠테이션을 만들어달라는 요청에 순식간에 텍스트와 이미지로 꾸며진 결과물을 만들어내는 장면은 충격적이었다. 그 이틀 전에 오픈AI는 이전 버전을 훌쩍 뛰어넘는 업그레이드 버전의 멀티모달 생성 AI 모델 ‘GPT-4’를 공개했다.

구글도 가만있지 않았다. 마이크로소프트의 각종 발표에 때를 맞춰 대화형 AI ‘바드’를 공개했고, 생산성 앱 제품군인 구글워크스페이스의 생성 AI 기능을 발표했다. 대규모 언어 모델 PALM의 활용 분야를 확장하는 행보도 이어갔다. 구글 검색 서비스에 생성 AI를 결합하겠다는 계획도 내놨다. 구글 바드는 공개 초기에 잘못된 답변을 내놔 내외부에서 많은 비판을 받았다. 이후 일반에 출시된 구글 바드는 프로그래밍 역량을 탑재한 업그레이드된 모습이었다.

이제 AI는 사람의 전유물 같던 창의적 사고의 영역도 넘보는 듯하다. 문학을 제외한 기타 사업적, 학문적 글쓰기는 AI의 우위 분야로 넘어갔다.

빌 게이츠는 “AI가 올해 가장 뜨거운 주제로 떠오를 것”이라며 “PC나 인터넷과 비슷한 영향을

사회에 미칠 수 있다”고 예측했다.

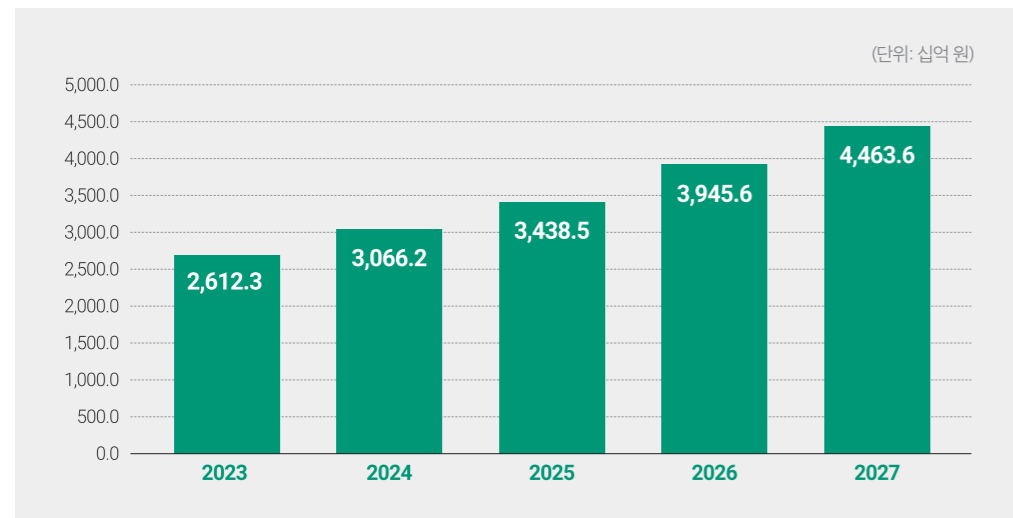
사회 전반에 미치는 AI의 영향력

골드만삭스는 최근 연구보고서에서 생성 AI의 활용으로 10년간 세계 국내총생산(GDP)을 7% 증가시킬 수 있다고 분석했다. 그러면서 정규직 일자리 3억 개가 AI 대체 위기에 노출될 것으로 예측했다. 일자리 3억 개는 전 세계 직업의 4분의 1에 달한다. 그 대상도 단순 노무직이 아니라 변호사, 세무사, 보험청구사 등 전문가 분야다. 생성 AI가 정보 검색, 문서 작성, 계산 등에서 사람보다 훨씬 더 효율적이기 때문이란 설명이다.

골드만삭스는 생성 AI 소프트웨어의 총 시장 규모를 1,500억 달러로 추산한다. 생성 AI는 비즈니스 워크플로우를 간소화하고, 일상적인 작업을 자동화하며 차세대 비즈니스 애플리케이션을 생성할 수 있는 것으로 기대된다. 비즈니스 애플리케이션에 침투한 생성 AI는 지식 근로자의 일상적인 효율성을 개선하고, 과학자나 소프트웨어 개발자의 연구 및 개발을 가속할 것이라고 전망했다. 더불어 몇몇 소프트웨어 회사는 이미 생성 AI로 제품 포트폴리오를 무장하고 있다고 강조했다.

골드만삭스 리서치는 “세계 경제와 사회에 생성 AI가 얼마나 많은 영향을 미칠지 알 수는 없고, 효과 발휘에도 시간이 걸릴 것”이라면서 “다만, 그 효과가 심오할 수 있다는 징후는 분명히 존재한다”고 밝혔다.

↓ 국내 AI 시장 전망(2023-2027년), 출처: IDC



한국IDC는 최근 발간한 ‘국내 인공지능 분석 시장 전망, 2023-2027’ 연구 보고서에서 국내 AI 시장이 올해 전년 대비 17.2% 성장해 2조 6,123억 원의 매출 규모를 형성할 것으로 전망했다. 국내 AI 시장은 향후 5년간 연평균 성장률 14.9%를 기록하며 2027년까지 4조 4,636억 원 규모에 이를 전망이다.

다양한 산업에서 AI 채택을 가속화하는 가운데, 디지털 기술과 산업 기술이 융합된 AI 생태계가 강화되고, 여러 서비스 사업 혁신이 이루어지면서 해당 시장 성장을 견인하는 것으로 나타났다. 더 많은 기업이 AI 관련 전문지식을 확보하기 위한 투자를 이어감에 따라 성장 추세가 지속될 것으로 전망된다.

보고서는 2023년부터 2027년까지의 국내 AI 시장의 주요 동향과 세부 항목별 시장을 전망했다. 해당 시장은 2027년까지 다양한 디지털 기술과 산업에서 사용되는 기술의 융합, 예측 및 추천 솔루션 고도화, 생성 AI를 통한 새로운 기술 활용 범위 확장 등의 변화를 유발할 것으로 보인다. 생성 AI와 비즈니스 생산성을 높여주는 자동화 솔루션이 AI 전체 시장의 성장세를 견인할 것으로 예상된다.

AI 기술의 명과 암

생성 AI의 잠재력에 긍정적 시각도 있지만, 비관적 전망도 많다. 특히 사람의 일자리가 AI로 대체될 것이란 두려움이다. 실제로 얼마전 IBM 아빈드 크리슈나 최고경영자(CEO)는 블룸버그와 인터뷰에서 향후 인사관리(HR) 등 백오피스 업무의 고용을 중단하거나 둔화시키고, 고객과 대면하지 않는 업무의 30%를 5년 안에 AI로 대체하겠다고 밝혔다. 올해 7,800명 규모의 구조조정을 단행하면서 그중 백오피스 업무를 사람 대신 AI로 바꾼다는 것이다. 그는 “고객을 대면하지 않는 회사 내 역할은 약 2만 6천 명의 근로자에게 해당한다”며 “5년 동안 그 중 30%는 AI와 자동화로 대체되는 것을 쉽게 볼 수 있을 것”이라고 말했다.

아예 생성 AI 개발을 잠시 중단하고 숨을 고르자는 주장까지 대대적으로 나왔다. 미국 비영리단체 미래생명연구소(FLI)는 ‘GPT-4 기능을 넘어서는 AI 개발을 6개월간 잠정 중단하자’는 공개 서한을 3월 28일 발표했다. 특히 FLI 지지자 명단에 일론 머스크 테슬라 최고경영자(CEO), 스티브 워즈니악 애플 공동창업자, 딥러닝 창시자인 요슈아 벤지오 몬트리올대 교수, 유발 하라리 등이 서명한 것으로 나와 주목을 받았다.

해당 서한은 '생성 AI가 인간 언어를 이해하고 구사하는 데 필요한 대규모 언어 모델을 구축하는 과정에서 데이터를 어떻게 처리하는지 분명하게 알 수 없다'고 지적했다. 또 생성 AI로 인한 허위정보 확산, 해킹, 대규모 실업 등에 대해서도 우려했다. 물론, 공개서한에 이름을 올린 인물 중 다수는 서명하지 않겠다고 정반대 입장을 밝히기도 했다. 확실한 건 미국 사회 전반에서 생성 AI에 대한 공포감과 반발이 작지 않다는 것이다.

생성 AI 기술을 활용하려는 산업계의 움직임은 전반적으로 일어나고 있다. 그동안 사내 연구실에서 소규모로 시도했던 AI 기술과 서비스들은 단기간 내 대외 고객을 위한 실제 서비스로 전환해야 하는 상황에 처했다. 생성 AI 모델의 학습과 추론에 필수 재료로 엔비디아의 GPU가 지목되며, 전 세계적으로 고성능 GPU 품귀 현상이 나타나고 있다.

생성 AI 기술을 얼마나 비용 효율적으로 활용할 것인가는 새로운 고민거리다. 세미애널리시스의 보고서에 따르면, 오픈AI의 챗GPT 운영 비용은 하루 70만 달러, 쿼리당 36센트로 추정된다. 챗GPT 작동에 고성능 서버가 필요하기 때문에 서버 운영비와 전력 사용료 역시 막대한 비용을 수반한다. 전문가들은 오픈AI의 서비스 상용화를 위해 올해에만 엔비디아 GPU를 3만 개 더 추가 주문해야 할 수도 있다고 예상한다.

생성 AI의 성공적 도입을 위한 선결 과제

메타의 비공개 AI 시스템 리버스 엔지니어링 담당 개발자인 필립 왕은 1월초 챗GPT와 유사한 기능을 할 수 있는 텍스트 생성 AI 모델 'PaLM+RLHF'를 깃허브에 오픈소스로 공개했다. 기술적으로 챗GPT 같은 서비스를 만드는 건 마음만 먹으면 가능한 일이란 걸 보여준 것이다.

그러나 문제는 비용이었다. AI 모델 학습에 필요한 컴퓨팅 자원과 데이터를 확보해야 하는데, 수백만 달러의 컴퓨팅과 데이터가 필요하다고 필립 왕은 밝혔다. PaLM의 매개변수 규모는 5,400억 개이며 챗GPT에 활용되는 GPT-3의 매개변수는 1,750억 개다. GPT-3 실행 비용은 AWS 인스턴스에서 연간 8만 7천 달러로 추정된다. 지난 2020년 한 연구에 의하면 매개변수 15억 개 규모의 텍스트 생성 모델을 개발하는 비용은 최대 160만 달러다. GPT-3 수준의 오픈소스 언어 모델인 '블룸(Bloom)'은 학습에만 엔비디아 A100 GPU 384개를 3개월 동안 돌려야 했다.

학습 비용은 생성 AI 모델 완성 후 사라질 수 있다. 그러나 생성 AI를 실제 프로덕션 환경에서 활용하려면 '추론 비용'이 발생한다. 그동안 예측 모델에 활용되는 AI 기술은 추론 시 상대적으로

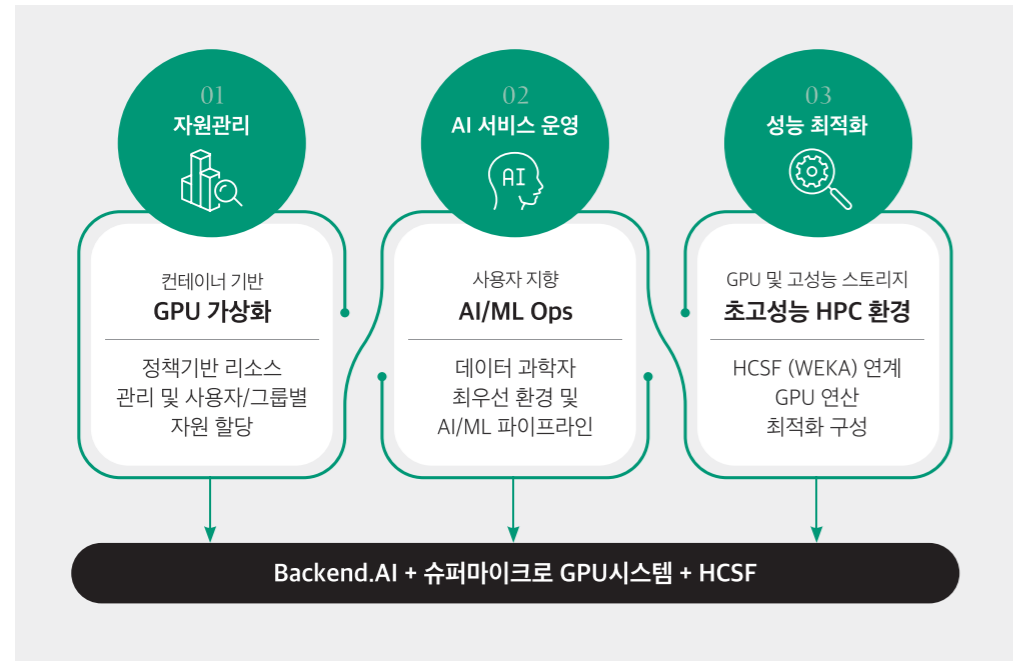
저렴하고 작은 용량의 인프라를 필요로 했다. 생성 AI는 학습에 준하는 규모의 추론용 인프라를 요구한다. 이 때문에 생성 AI로 진정한 비즈니스 성공을 구가하려면 모델 운영 비용과 서비스 수익 사이의 균형을 찾아야 한다는 결론에 도달하게 된다.

현재 업계는 AI 추론 전용 반도체인 'NPU(Neural Processing Unit)'에 기대를 걸고 있다. 구글은 TPU를 생성 AI 추론 비용 절감에 활용할 뜻을 밝히고 있고, 마이크로소프트도 자체 AI 칩인 '아테나(Athena)'를 개발해 애저와 오픈AI에서 테스트 중인 것으로 알려졌다. 국내에도 NPU 개발에 도전장을 던진 퓨리오사AI, 사피온, 리벨리온 등의 기업이 있다. NPU를 활용하면 GPU의 25% 비용으로 AI 학습과 추론이 가능할 것으로 예상된다.

NPU 개발은 아직 완성 단계에 이르지 못했다. 반면, 기업은 당장 AI 서비스를 개발하고 운영해야 하는 상황이다. 비싼 도입 비용과 운영 비용, AI 관련 기술의 난해함, 부족한 기술 인력 등은 선뜻 AI 세계로 나아가도록 결심하기 힘든 요소다.



↓ 효성인포메이션시스템이 제공하는 'AI 플랫폼'



AI 플랫폼으로 최적의 환경 구축

당면과제로 떠오른 AI 도입 문제를 해결하고, 자사에 적합한 AI를 도입하고 운영하기 위해 '통합 AI 플랫폼'이 대안으로 떠오르고 있다. 통합 AI 플랫폼은 GPU 서버와 고성능 병렬 파일 스토리지, 네트워크, 컨테이너 등의 솔루션을 결합해 사전에 통합된 플랫폼 형태로 제공된다. 연산과 데이터 저장에 필요한 하드웨어와 컨테이너 기반 GPU 가상화 운영 솔루션을 바로 활용할 수 있으며, AI 시작부터 구축, 운영에 이르기까지 단일 벤더와 소통하며 프로젝트를 진행할 수 있다.

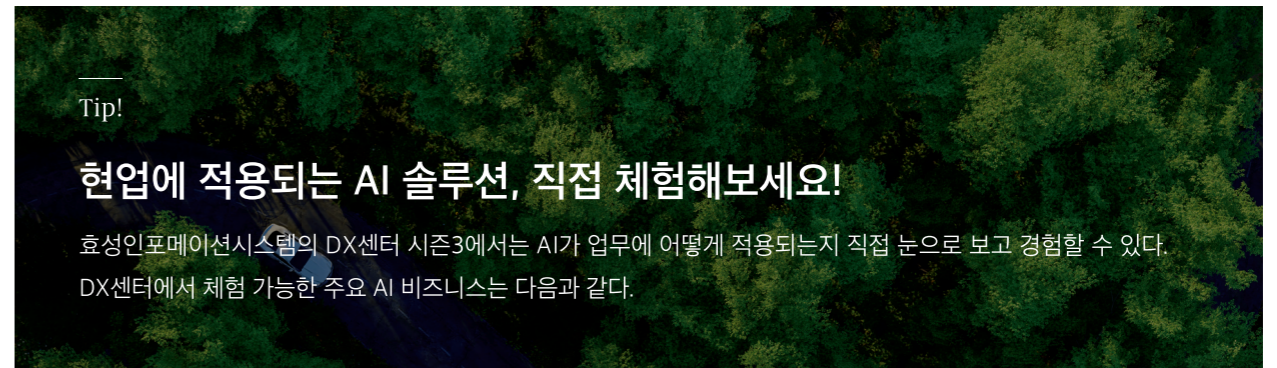
효성인포메이션시스템은 다양한 유형의 업무에서 AI를 적용할 수 있도록 고객 맞춤형, 프로젝트 맞춤형으로 통합 AI 플랫폼을 제공한다.

새로 AI를 도입하려는 기업은 AI 도입 후 성과와 방향성을 확신하기 힘든데, 최소 비용으로 AI 플랫폼 환경을 경험해 볼 수 있는 AI 테스트베드 환경을 제공한다. 이미 보유한 GPU와 각종 인프라 자원을 재활용하는 한편, 새로 도입해야 할 인프라 자원의 도입 계획까지 추정해 준다. 효성인포메이션시스템의 통합 AI 플랫폼 풀 패키지를 이용하면 AI 개발에 필요한 학습 시간을 최소화하고, 운영과 비용 효율을 높이는 데 도움을 받을 수 있다.

효성인포메이션시스템은 통합 AI 플랫폼을 통해 전체적인 AI 인프라 구성부터 최적화 기술을 기반으로 자문, 컨설팅 및 계획, 설계, 구축, 수행까지 단일 벤더에서 통합 제안한다.

미래는 어떤 모습일까? 우리는 어떤 세상에서 살게 될까? IT 기술의 발전은 우리의 삶과 산업에 어떤 변화를 가져올까? 지난 3개월 동안 세상에서 가장 많이 던진 질문일 것이다. 그리고 공포를 느끼는 반 AI 세력의 등장은 어찌 보면 당연하다.

그렇다고 생성AI와 IT 기술의 발전을 인위적으로 막을 수 있을까? 답은 '불가능'이다. FLI 성명 발표 직후 유수의 AI 전문가와 IT 종사자들이 필연적 흐름을 억지로 막으려는 시도라며 비판한 것만 봐도 알 수 있다. 오히려 흐름을 읽고, 그 내용을 살핀 뒤 미래에 어떻게 대응할지 면밀히 따지는 게 현명한 선택일 것이다.



Tip!

현업에 적용되는 AI 솔루션, 직접 체험해보세요!

효성인포메이션시스템의 DX센터 시즌3에서는 AI가 업무에 어떻게 적용되는지 직접 눈으로 보고 경험할 수 있다. DX센터에서 체험 가능한 주요 AI 비즈니스는 다음과 같다.

01 효율적 GPU 자원관리

기존 GPU 자원 관리의 한계를 극복하기 위한 컨테이너 기반의 GPU 가상화로, 사용자 및 그룹별 정책 설정에 의한 GPU 자원관리 방안을 소개한다.

02 AI/ML Ops 솔루션

표준 웹 GUI 환경에서 사전 정의 개발 환경(Tensorflow, Pytorch 프레임워크, CUDA) 등을 구현해 볼 수 있다. 데이터 수집, 가공, 모델개발, 운영 배포의 AI/ML 오픈스 파이프라인 가이드를 체험해 볼 수 있으며, 특히 AI 모델 학습 및 추론 시연도 가능하다.

03 성능 최적화를 위한 AI 인프라 설계 가이드

초고성능 스토리지, 고속 네트워크, GPUDirect, NV링크 및 NV스위치를 지원하는 GPU 서버 구성을 통해 성능 최적화 가이드를 제안한다. GPUDirect 스토리지와 NV링크 성능 테스트 결과 분석에 따라 기업별 인프라 도입 방향을 제시하고, 기업의 환경 및 고객의 요구사항에 맞춰 효성인포메이션시스템이 제안하는 AI 플랫폼 도입 가이드를 만날 수 있다.