미국에서 '데이터 사이언스' 전공자의 설자리가 점차 없어진다는 이야기. 요약하면,

.

1. 7-8년 전 AI 붐을 타고 데이터 과학이라는 분야가 유망해졌고 관련 일자리도 급증.

.

2. 하지만 그동안 AI 기술 자체가 발전하고 데이터 관리가 자동화되면서 점차 사람의 손을 덜 타고 있음. 단순직은 사라지고, 도메인 지식, 수학, AI 기술 등 좀 더 심화된 역량까지 가진 인력들만 필요해짐.

.

3. 애초에 '데이터 과학'은 역사가 짧고 '근본'이 없는 학문임. 이제 '데이터 과학자'가 되고 싶다면, 데이터 과학을 전공하면 안 됨. 통계, 산공 등 '근본'있는 학문을 전공한 뒤, 데이터 과학은 스킬 셋으로 접근해야 함.

.

.

2010년대 중반 AI 붐이 일었을 때, AI는 곧 딥러닝이고, 딥러닝은 곧 신경망이었다. 그리고 신경망을 '잘' 훈련시키려면 '양질'의 데이터가 필요했다. 의도된 대로 신경망이 잘 동작하려면, 해당 도메인을 대표하는 데이터를 선별해 주입시켜줘야 하기 때문이다.

.

비유가 적절한지 모르겠지만, 아이를 '손흥민'같은 1류 선수로 키우려는데, 문학, 바이올린, 역사, 철학, 야구를 학습시켜 봐야 '축구'에 하등의 도움이 안 된다. 유연성, 균형감, 동체시력, 반응속도, 체력 등 축구에 관련된 스킬 셋만 집중적으로 학습시켜야 한다 (물론 '손흥민'을 복제하듯 학습시킬 수도 있지만, 이 경우 조건이 조금만 달라지면 아무것도 못하는 반쪽짜리 선수가 된다. 이른바 Overfitting).

.

따라서, 일단 데이터를 다루는 사람은 숫자를 보고 해석할 줄 아는 능력, 즉, 위의 예에서의 '축구'라는 도메인 지식이 있어야 한다. 이런 배경을 바탕으로 다양한 소스들로부터 채취된 데이터를 해석, 노이즈 제거, 인사이트 추출, 관리, 가시화할 수 있는 전담인력에 대한 수요가 높아지게 된 것이다. 데이터를 체계적으로 관리하는 툴, 스크립트 언어들도 등장하면서 관련 산업의 전망이 밝아졌다.

.

덕분에 AI를 핵심 기술로 사용하는 기업마다 별도의 데이터 부서가 조직되었고, 관련 인력들이 충원되기 시작했다. 유사한 일을 하던 데이터 베이스 전공자, 수학자, 통계학자들이 발 빠르게 이 분야로 진입했다. 물들어올 때 노 젓듯, 미국 각 대학에서는 데이터 사이언스 석사과정을 우후죽순 개설하기 시작했다.

.

그런데 시간이 가면서 상황이 역전되기 시작했다. 그동안 '데이터 과학'은 AI가 일을 잘할 수 있도록 도와왔는데, 이 배은망덕한 AI는 자신이 좀 똑똑해졌다고 데이터 과학을 괄시하기 시작한 것이다. "내가 니 일까지 할 수 있으니, 너는 더 이상 필요 없어"라고. ChatGPT가 그 방점을 찍었다. ChatGPT과 연동되는 다양한 통계 플러그인들이 데이터 과학자들의 일을 상당 부분 자동화시켜 버린 것이다.

.

문제는 뒤늦게 미국에 건너와 '데이터 과학'을 전공한 유학생들의 졸업 시점과 이러한 상황이 맞물린 것이다. 현재 유학생들은 인턴을, 졸업생들은 관련 일자리를 쉽게 구하지 못하고 있다. 벌써 업계는 데이터 관련한 단순 직무를 자동화시키고 있고, 남은 자리는 실무 경험이 많고, 대규모 데이터를 다뤄본 경력자에게만 돌아가고 있기 때문이다.

.

그 원인은 아래 비디오 출연자가 말했 듯 '데이터 과학'이란 '근본'이 없는 학문이기 때문이다. 근본이 없다고 하니 관련 전공자들은 '내가 데이터 과학 공부하느라 코피 쏟은 시간, 돈이 얼만데'하며 발끈하실 수도 있다.

.

여기서 내가 '근본'이 없다고 한 이유는 데이터 과학이 많은 전공과 얽혀있기 때문이다. 수학, 통계학, 컴퓨터 공학, 산업 공학 등 기존에 오랜 역사를 가진 '근본'있는 전공에서 필요한 과목들을 취사 선택한 학문이라는 말이다.

.

관련 산업이 성장하고 인력의 수요가 늘어나자, 해당 직무에 특화된 전공을 급조했다. 필요한 과목들은 기존의 여러 전공들에 산개되어 있었고, 이들 중 체리피킹해 그럴듯한 이름의 학문으로 재창조한 것이다. 좋게 말하면 '융복합' 학문인 셈. 하지만 늘 그렇듯 융복합은 깊이만 얕아지는 부작용을 낳는다.

.

즉, 제대로 하려면 한도 끝도 없이 학제 간 전공을 모두 공부해야 되고, 취사 선택된 과목만 공부하면 전문성을 잃는 것이 이 '데이터 과학'이라는 학문인 것이다. '시스템 반도체'가 미래의 먹거리라며 한국에서 반도체 학과를 신설하는 것과 같은 이치다.

.

덕분에 현재 미국에서 꽤나 오남용 되고 있는 타이틀이 바로 '데이터 사이언티스트'다. 근본 있는 자신만의 전문성을 바탕으로 데이터를 다뤄본 오랜 실무 경험을 가진 이들 못지않게, 부트캠프에서 얕게 통계, ML 스킬 셋을 공부하고 나온 이들도 '데이터 사이언티스트'라고 자처하고 있기 때문이다.

.

따라서 트렌드에 민감한 '학문'을 전공으로 선택하는 것은 매우 주의해야 한다. 고용 시장에서 수요가 많은 시점엔, 미리부터 '저점 매수'한 사람들만이 그 열매들 딴다. 수요가 많으면 관련 학과가 신설되는 등 자연스럽게 공급도 많아지는데, 시류에 민감한 분야는 그만큼 수요도 빨리 줄어들기 마련이다. 이런 학문을 전공한 사람들은 줄어든 자리를 두고 시장에 쏟아지는 인력들과 경쟁해야 한다.

.

개인적으로 AI/ML 엔지니어도 안전하지 않다고 본다. 알려진 신경망들을 조합한 뒤, 자동화로 생성된 데이터들을 반복적으로 돌려보고, 그 결과를 테스트 부서에 넘겨주는 일은 그다음으로 자동화될 타깃이기 때문이다. 'AI 기술 그 자체'를 연구하는 극소수의 엔지니어/연구자를 제외하고, 깊이가 얕은 업무를 하는 ML 엔지니어는 점차 그 자리가 줄어들 것이다.

.

AI가 발전하면 발전할수록, 그 기술이 광범위한 분야로 전파될 것인데 이는 AI가 보편화된다는 의미다. 이럴 때일수록 더 '근본'있는 학문을 전공해야 한다. 미래에는 자신만의 도메인 전문성을 가지고 AI를 '생산성 도구'로 활용하는 자만이 살아남을 것이기 때문이다.

Power bi / pytorch / 7zip / git / notepad/ sourcetree /

VMware

Spyder – python

Sqlite

Mysql

Star UML

R-Studio

Ubuntu

Java